DOCUMENT RESUME

ED 226 611 FL 013 521

AUTHOR Yates, James; And Others
TITLE Baker DeKanter Review: Inappropriate Conclusions on the Efficacy of Bilingual Education.
PUB DATE Dec 82
NOTE 192p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Viewpoints (120)

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS *Bilingual Education; Educational Policy; *Educational Research; Language of Instruction; Research Design; Research Methodology; *Research Problems; Research Reports

ABSTRACT

Methodological problems associated with the research cited by the Baker DeKanter Review of Bilingual Education (1981) are reported. The review concluded that there is little evidence supporting the need for bilingual instruction for language minority children. Given the widespread legal, fiscal, policy, and media implications of this conclusion, it is important that the evidence be analyzed. Each of the 28 studies cited to support the Baker DeKanter conclusion was analyzed in depth in regard to methodological soundness. Numerous instances of the following methodological errors were found: non-random assignment of subjects, high attrition of subjects, extremely small sample sizes, inappropriate measurement instruments or procedures, inappropriate pretest/posttest time frames, inconsistent design, lack of control of critical learning variables, variations in qualifications of instructional personnel, and lack of recency of the studies cited. Given these problems, the conclusions of the Baker DeKanter review should be questioned. Detailed reviews of each of the 28 studies are appended. (RW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED226611

BAKER DEKANTER REVIEW:

INAPPROPRIATE CONCLUSIONS ON THE EFFICACY OF BILINGUAL EDUCATION

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James Yates

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

James Yates
James Adamko
Linda Avila
Sandra Burks
Shernaz Garcia
Betty Horton
Pamela McCollum
Emily Poling
Michael Thomas
John Westbrook

FL 013 521

THE UNIVERSITY OF TEXAS AT AUSTIN

DECEMBER, 1982

Baker/DeKanter Review:

Inappropriate Conclusions on the Efficacy of Bilingual Education

The 1980 decennial census placed the number of Hispanics in the United States at 14.6 million based upon self-identification according to Spanish origin. Popular publications, such as Time and Newsweek, estimate the Hispanic population at closer to 25 million. From 1970 to 1980, the Hispanic population of the United States grew at a rate 6.5 times greater than the general population. The growing Hispanic community in this country emphasizes the growing political, economic, and international importance of Hispanics. It is of extreme significance, therefore, given the fact that the majority of Hispanics in this country are within the "school age" population, that all citizens, particularly educators, be concerned with the efficacy of schooling being provided this ethnic group.

On September 25, 1981, a final draft report was completed by staff members in the Office of Planning, Budget, and Evaluation of the U.S. Department of Education. That report was titled "Effectiveness of Bilingual Education: A Review of the Literature." Authored by Keith A. Baker and Adriana A. DeKanter, this report is frequently referred to in the literature, and in the popular press, as the Baker/DeKanter Review of Bilingual Education. Baker and DeKanter focused their review upon what they described as twenty-eight studies "found applicable" with "outcomes about which we can be reasonably confident," and "found to apply to our concerns and to meet our methodological criteria." Baker and DeKanter, in their report, go to great lengths with a substantial number of introductory comments, to build an aura of careful scholarship surrounding their review, of a clear and well-defined conceptual basis for their review of specific studies, of a comprehensive and extensive selection of literature, and of conclusions and recommendations based upon clear evidence derived from their data bases.

Baker and DeKanter (1981) conclude "The case for the effectiveness of transitional bilingual education is so weak that exclusive reliance on this instructional method is clearly not justified." (p. 1 abstract) "There is no justification for assuming that it is necessary to teach non-language subjects in the child's native tongue in order for the language-minority child to make satisfactory progress in school." (p. 1 abstract)

There have been very few literature reviews that have required or justified a careful and specific critique; however, the importance of such a critique of the Baker/DeKanter Review can be determined by the following evidence:

1) Policy and Fiscal Effects. Senator Walter Huddleston, the Honorable Senator from Kentucky, introduced on December 16, 1981, Senate Bill 2,002. This piece of legislation would allow bilingual education to be provided for one year only, with the possibility of an additional year of bilingual instruction if a child has had an extensive individual evaluation which establishes the need for continued services. But, in no event would

the child be permitted to be enrolled in bilingual education in excess of three years. Senator Huddleston states in prepared remarks for the Congressional Record of Wednesday, December 16, 1981, that a major rationale for the introduction of Senate Bill 2,002 is the information provided by "a recent report by the Department of Education...conclusively shows that the bilingual education program is not working... The report concluded that the case for the effectiveness of transitional bilingual education is so weak that exclusive reliance on this instruction method is clearly not justified." (Huddleston, 1981, No. 188, Part 3) The report to which Senator Huddleston refers is the Baker/DeKanter Review.

The Federal Office of Budget and Management has recommended significant cuts or reductions in the level of federal funding for bilingual education. One of the bases for these recommended reductions, according to Foster (1981), is the conclusions described in the Baker/DeKanter Review of the Effectiveness of Bilingual Education.

2) Legal. Various legal action has been initiated based upon the information included in the Baker/DeKanter Review. For example, Robinson (1981) states "The Attorney General hopes to use the Education Department's study to support the state's claim that bilingual education as called for by Judge Justice (U.S. vs. Texas,

1981) is not necessarily the most effective way to remedy language problems of Spanish speaking youngsters." (p.5).

3) Research Procedures and Scholarship. The Baker/DeKanter Review has become the basis of and formulation for certain research associated with bilingual education. Foster and Matske (1982) indicated that a major and expensive ($500,000) study funded by the Office of Education to a private research corporation (RMC Research Corporation) based one of two major sections of conclusions upon the Baker/DeKanter Review. That particular study was to look primarily at Department of Education, Title VII funded, bilingual education programs. A scholarly review of the Baker/DeKanter Report was recently completed by Seidner (1981). Seidner openly questions many of the conclusions, methodologies and so forth that were utilized by Baker/DeKanter.

4) Conceptual, Philosophical, Capital Development. The Baker/DeKanter Review has begun to serve as the basis for the development of philosophical or conceptual postures relative to the education of children of limited English proficiency (Campbell & Gray, 1982; Chavez-Hernandez, et.al.,1981; Robledo, 1982.)

5) Media Dissemination. The Baker/DeKanter Review has served as the basis of a substantial number of

> newspaper, magazine, television, and other news media information. In almost all instances, media portrayed the Baker/DeKanter Review as substantive evidence of the lack of efficacy of bilingual education (Associated Press, 1981; Corpus Christi Caller, 1981; Real, 1981).

Given the multiple significant effects of the dissemination of the Baker/DeKanter Review, it seems important, if not essential, that a scholarly inquiry be made into the specific studies cited by Baker and DeKanter as "methodologically applicable." Therefore, a research team* was assembled at the University of Texas at Austin for the purpose of completing such a scholarly inquiry.

The purpose of this paper is to report one small aspect of the work of a research team at the University of Texas. An appendix to the paper is provided for those scholars who would like to analyze in depth the reviews completed by the research team. Each review in the appendix includes careful references and documentation to the original source so that any concerns with the review can be followed to the original document. This paper will provide a relatively brief reporting of specific problems associated with the procedures and methodologies of the "twenty-eight" studies selected by Baker and DeKanter as the only studies that were "methodologically applicable."

---

*Comprised of the following members: James Adamko, Linda Avila, Sandra Burks, Shernaz Garcia, Betty Horton, Pamela McCollum, Emily Poling, Michael Thomas, John Westbrook, and James Yates.

A range of procedural reporting difficulties are present in the Baker/DeKanter volume itself, such as poorly identified publication/references, hasty writing and/or editing as exemplified by grammatical and typographical errors, difficulties in pagination, narrative that is discrepant with tables and so forth. These problems would, in and of themselves, allow one to deduce that there are difficulties with the Baker/DeKanter Review. However, these difficulties, in and of themselves, would not necessarily lead to difficult conclusions.

In order to control possible potential researcher prejudice which could develop from seeing such lack of care and scholarship, the research team at the University of Texas at Austin did not read the Baker/DeKanter report prior to initiating their review efforts of the "twenty-eight" selected studies. Procedurally, the University of Texas research team took the "twenty-eight" references from Baker/DeKanter and sought the original source, that is to say, the journal article, project report, paper presented, and so forth. This, in itself, was an extremely time-consuming and difficult task, as, to say the least, many of the citations are most obscure literature. Citations were often generally unavailable and required the most tenacious efforts of inquiry on the part of the researchers in order to find the original documents.

The second procedure for the University of Texas research team was to utilize the same categories and topic headings for their review or critique of the specific studies as were utilized by Baker and DeKanter.

The following difficulties discovered by the University of Texas research team reflect only those most striking errors that are clearly evident and documentable in the original sources. The errors described do not reflect any effort at deduction or interpretation on the part of the research team. The following section reports those errors in terms of number of times the error was noted in the original sources, and provides specific examples of the errors.

1. Non-random assignment of subjects and/or other procedures which would result in non-equivalent groups - 21 instances.

   Examples: Carsrud & Curtis (1980) used no bilingual or Spanish dominant students in the control groups; only English dominant students were placed in control groups.

   Mathews (1979), students with lowest fluency were assigned to treatment groups.

   Lum (1971), assignment to treatment and control was based upon area of residency and availability of bilingual classes, rather than randomization.

2. High attrition of subjects - 5 instances.

   Examples: Stebbins (1977), 70% of treatment and control group students were not present for post-testing.

   Ramos, Aguilar & Sibayan (1967), 758 treatment and 1,164 control students diminished to 232 treatment and 301 control students at the time of post-testing.

   Balasubramonian et. al. (1973), 26% attrition occurred in the treatment group and 21% attrition occurred in the control group.

3. Discrepancy in size of n or very small n - 8 instances.

   Examples: Barik & Swain (1975) used as few as ten randomly selected students for some cohort groups.

   Legarrata (1979) had an n of seven in one treatment group.

   Kaufman (1968) had 30% more treatment subjects than control subjects.

4. Inappropriate measurement instruments and/or procedures - 26 instances.

   Examples: Barik, Swain & Nwanunobi (1977) measured reading achievement in the second grade on instruments requiring reading, yet, reading according to the school curriculum, was not introduced until the end of the second grade.

   Legarrata (1979) used cookies, peanuts, and raisins as reinforcement during timed tests of oral language proficiency. Additionally, Legarrata used four 11- and 12-year old girls as examiners to collect language proficiency data. Two of the girls were the investigator's daughters.

   Lum (1971) administered an instrument for the purpose of determining mono/biculturalism. In actuality, the test was a measure of language usage.

   Campeau, Roberts, Bowers, Austin & Roberts (1975) did no screening for eligibility. Participants volunteered.

   Danoff (1977a, 1977b, 1978a, 1978b) teacher judgment and perception was the only measure of language dominance.

5. Time frame of pre/post-testing inappropriate - 8 instances.

   Balasubramonian et. al. (1973), only five months elapsed between pre and post-testing.

Carsrud & Curtis (1980) only five to six months pre to post-testing occurred.

Kaufman (1968), had wide differences in the time frame of pre/post-testing: School A - 18 months pre/post-testing; School B - 9 months.

Moore & Parr (1978) provided six to eight months pre/post-testing.

6. Inconsistent design implementation/inconsistent treatment - 35 instances.

Examples: Lambert & Tucker (1972) in kindergarten and first grade an immersion model was utilized as treatment, and in second through fourth grades, a partial bilingual model was utilized as treatment. Yet, results are cited as effects of an immersion model.

Zirkel (1972), treatment groups varied in amount of bilingual instruction from ten minutes to one hour with individual versus group instruction, tutoring versus self-contained classrooms; all instruction was grouped into one treatment group.

Ames & Beck (1978), during the middle of the year of the study, the School Board mandated a reduction of native language instruction to 50% of the time for the treatment groups.

Carsrud & Curtis (1980) teachers in the bilingual classroom treatment groups taught 82.5% of the time in English only. One third of the treatment group teachers taught only in English.

Stern (1975) at the end of the project, it was discovered the

the highest scorers in the control group had at one time been students in the treatment group.

Campeau, Roberts, Bowers, Austin & Roberts, (1975) treatment groups received instruction in a variety of settings: self-contained, open space, team teaching, and so forth, yet were analyzed as one treatment.

7. Lack of control of known critical learning variables

10 instances.

Examples: A variety of studies - Legaratta (1979), Kaufman (1968), Moore & Parr (1978), McSpadden (1981), did not control for varying intelligence in treatment and control groups. Huzar (1973), Covey (1973), Mathews (1979), did not control for different socio-economic levels of treatment and control groups. Time on task was not controlled in a variety of studies, such as McSpadden (1981), Mathews (1979).

8. Different standards/qualifications of instructional personnel

14 instances.

Examples: Balasubramonian et. al. (1973), control group teachers had twice as many years of experience as the treatment group teachers.

Campeau, Roberts, Bowers, Austin & Roberts (1975) control group teachers had an average of ten years experience, treatment group teachers were all first and second year teachers.

Skoczlas (1972), aides without teaching credentials had equal responsibility for instruction with certified teachers.

Stern (1975), teachers with no certification or training as bilingual educators taught in the treatment bilingual education program groups.

Danoff (1977a, 1977b, 1978a, 1978b), teachers in the treatment Title VII classrooms widely varied in proficiency in both languages of interest.

9. Lack of recency of studies cited. 25% of the studies were ten or more years old; 40% of the studies were more than five years old.

In reality, there are not twenty-eight studies, as previously referred to by Baker and DeKanter, but twenty-six, as Olesini (1971) was counted as one of the twenty-eight acceptable studies, yet this study was removed by Baker and DeKanter into the section of the report which they deemed as unacceptable or inappropriate to their effort. Of interesting note, Olesini (1971) is a study which showed significant advantage for bilingual instruction with these conclusions limited primarily by a short test/retest period of only seven months. Additionally, Baker and DeKanter separately cite McSpadden 1979 and McSpadden 1981, when in reality, these two citations are one study, as the 1979 citation is a preliminary or interim report, with the 1981 citation being the final report of the same project.

Given the significant difficulties found with the studies that were selected by Baker and DeKanter, one must seriously question the conclusions reached by Baker and DeKanter, i.e., "the case for the effectiveness of transitional bilingual education is so weak that exclusive reliance on this instruction method is clearly not justified." (p.1 abstract)

Whatever may be the reason for the conclusions reached by Baker and DeKanter, it is now possible, based upon the efforts of the research team at the University of Texas at Austin, to allow scholars to examine the original sources, to examine the information obtained and produced by the research team, and to allow it to be scrutinized in a true atmosphere of open scholarship. Such reviews are not only acceptable to the research team, but are welcomed.

Those concerned with honesty of inquiry and scholarship must be informed and able to respond to the Baker/DeKanter document. Such knowledge is essential in order that appropriate policy decisions, resource allocations, fairness and equity of service delivery can occur, void of ignorance, prejudice, politics, and racism.

References

Ames, J. S. & Bicks, P. An evaluation of the Title VII bilingual-bicultural programs, 1977-1978 school year, final report. Community School District 22, Brooklyn, New York: July, 1978.

Baker, K. & DeKanter, A. Effectiveness of bilingual education: a review of the literature. Washington, D.C.: Department of Education, Office of Planning, Budget, and Evaluation, Sept. 25, 1981.

Balasubramonian, K., Seeley, H.N. & De Weffer, R.E. Do bilingual education programs inhibit English language achievement? a report on an Illinois experiment. Paper presented at the Seventh Annual Convention, Teachers of English to Speakers of Other Languages. San Juan, Puerto Rico, May 9-13, 1973.

Barik, H. C. & Swain, M. Three year evaluation of a large scale early grade French immersion program: the Ottawa study. Language Learning, 1975, 25(1) 1-30.

Barik, H. C., Swain, M., and Nwanunobi, E. A. English-French bilingual education: the Elgin study through grade 5. Canadian Modern Language Review, 1977, V(33), 459-475.

Campbell, R.N. & Gray, T.C. Critique of the U.S. Department of Education report on effectiveness of bilingual education: a review of the literature. IDRA Newsletter, January, 1982, p. 3-4.

Campeau, P.L., Roberts, A.I.H., Bowers, J.E., Austin, M. & Roberts, S.J. The identification and description of exemplary bilingual education programs. Palo Alto, California: American Institute for Research, August 1975 (Corpus Christi component of report, D 57-76).

Carsrud, K. & Curtis, J. Final technical report: ESEA Title VII bilingual project. Austin: Austin Independent School District, 1980, (Publication No. 79.21).

Chavez-Hernandez, E.; Llanes, J.; Alvarez, R. & Arvizu, S.F. The federal policy toward language in education: pendulum or progress. Sacramento, CA: California State University, Department of Anthropology, Cross-Cultural Resource Center, Monograph No. 12, 1981.

Cohen, A.D., A Sociolinguistic Approach to Bilingual Education: Experiments in the American Southwest. Rowley, Massachusetts: Newbury House, 1975.

Bilingual education. Corpus Christi Caller, October 5, 1981, p. 12A.

Education Department memo slams bilingual education. Corpus Christi Caller, Sept. 30, 1981.

Cottrell, M. C. Bilingual education in San Juan County, Utah: a cross-cultural emphasis. Paper presented at the American Education Research Association Annual Convention. New York, New York, February 4-7, 1971.

Covey, D. D. An analytical study of secondary freshmen bilingual education and its effect on academic achievement and attitude of

Mexican-American students. Unpublished doctoral dissertation, Arizona State University, 1973.

Danoff, M. N. *Evaluation of the impact of ESEA Title VII Spanish/English bilingual education programs. Volume I: study design and interim findings*. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, 1977.

Danoff, M. N. *Evaluation of the impact of ESEA Title VII Spanish/English bilingual education programs. Volume II: project descriptions*. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, 1977.

Danoff, M. N. *Evaluation of the impact of ESEA Title VII Spanish/English bilingual education programs. Volume III: year two impact data*. Palo Alto, California, American Institutes for Research in the Behavioral Sciences, 1978.

Danoff, M. N. *Evaluation of the impact of ESEA Title VII Spanish/English bilingual education programs. Volume IV: overview of study and findings*. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, 1978.

Foster, S.G. & Matske, M. U.S. study reports shortage of qualified bilingual education teachers. *Education Week*, June 9, 1982, p. 6.

Foster, S.G. Studies stirred debate over worth of federal programs. *Education Week*, *1*(15), December 21, 1981, p. 1.

Huddleston, W.D. *Congressional Record*, 97th Congress, Vol. 127, Wednesday, December 16, 1981, No. 188, Part 3.

Huzar, H. *The effects of an English-Spanish primary-grade reading program on second and third grade students*. Unpublished thesis, Rutgers University, May, 1973.

Kaufman, M. Will instruction in reading Spanish affect ability in reading English? *Journal of Reading*, 1968, *17*, 521-527.

Lambert, W.E. & Tucker, G.R. *Bilingual education of children: the St. Lambert experiment*. Rowley, Massachusetts: Newbury House, 1972.

Legarrata, D. . The effects of program models on language acquisition by Spanish-speaking children. *TESOL Quarterly*, 1979, *13*(4), 521-534.

Lum, J. B. *An effectiveness study of English as a second language*. Unpublished doctoral dissertation, University of California at Berkeley, 1971.

Mathews, T. *An investigation of the effects of background characteristics and special language services on the reading achievement and English fluency of bilingual students*. Seatle, Washington: Seatle Public Schools, November, 1979.

McConnell, B.B. *Effectiveness of individualized bilingual instruction for migrant students*. Unpublished doctoral dissertation, Washington State University, 1980.

McSpadden, J. R. *Acadiana bilingual-bicultural educational program: final evaluation report, 1980-81*. Lafayette, Louisiana: Lafayette Parish School Board, 1981.

Moore, F.B. & Parr, G.D. Models of bilingual education: comparisons of effectiveness. *Elementary School Journal*, *1978, 1979*(2), pp. 93-97.

Olesini, J. *The effect of bilingual instruction on the achievement of elementary pupils*. Unpublished doctoral dissertation, East Texas State University, 1971.

Pena-Hughes, Eva & Solis, Juan. *abcs*. McAllen, Texas: McAllen Independent School District, Mimeographed, 1980.

Plante, A.J. *A study of the effectiveness of the Connecticut 'pairing' model of bilingual-bicultural education*. Hamden, Connecticut: Connecticut Staff Development Cooperative, 1976.

Ramos, M., Aguilar, J. & Sibayan, B. *The determination and implementation of language policy*. Quezon City, Republic of the Philippines: Phoenix Press, 1967.

Real, D. Study claims bilingual education doesn't work. *San Antonio Light*, Sept. 30, 1981.

Robinson, V. Texas, U.S. Education Department spar over top-secret bilingual education report. *Education Times*, Sept. 21, 1981, p. 5.

Robledo, M.R. Bilingual evaluation research: does it work? *IDRA Newsletter*, February, 1982, pp. 1-7.

Seidner, S.S. *Political experience or educational research: an analysis of Baker and De Kanter's review of the literature of bilingual education*. Rosslyn, VA: National Clearinghouse for Bilingual Education, 1981.

Skoczlas, R. V. *An evaluation of some cognitive and affective aspects of a Spanish-English bilingual education program*. Unpublished doctoral dissertation, University of New Mexico, August, 1972.

Stebbins, L.B., St. Pierre, R.G., Player, E.C., Anderson, R.B., Crum, T.R., *Education as experimentation: a planned variation model. volume IV-A and IV-D, evaluation of follow through*. Cambridge, Massachusetts, ABT Associates, 1977.

Stern, C. *Final report of the Compton unified school district's Title VII bilingual-bicultural project: September, 1969 through June, 1975*. Compton, California: Compton Unified School District, September, 1975.

U.S. v. Texas, Civil Action 5281, 5th Circuit, U.S. District Court, Tyler Division, January 9, 1981.

Zirkel, P. A. *An evaluation of the effectiveness of selected experimental bilingual education programs in Connecticut*. West Hartford, Connecticut: Hartford University, Connecticut Migratory Children's Program, May, 1972.

Reference:	Ames, J. S. & Bicks, P. An evaluation of the Title VII bilingual-bicultural programs, 1977-1978 school year, final report. Community School District 22, Brooklyn, New York: July, 1978. Baker de Kanter Review - p. 22-23, Chapter 2.
Available From:	ERIC Document Reproduction Service, ED 185 132.
Name of Study:	An Evaluation of the Title VII Bilingual-Bicultural Programs, 1977-1978 School Year, Final Report.
Author and Date:	Ames, J.S. & Bicks, P.; 1978.
Location:	Community School District 22, Brooklyn, New York.
Treatment Group:	212 Spanish and French Creole-speaking students in Grades 1-9 of Community School District 22, Brooklyn comprised the treatment group. There were seven bilingual classes located in four buildings for Grades 1-9. Two of these were bilingual French-English classes (Grades 1-6). Additionally, intensive ESL classes were conducted at one elementary and two junior high schools.
Control Group:	457 students who were identified as language deficient and were served through regular classes with "pull-out" programs. Each student received 3-5 hours per week of intensive ESL instruction.
Duration:	1977-1978 school year.
Ages:	Grades 1-9.
Type of Program:	Spanish and French Creole transitional bilingual-bicultural education and intensive ESL.

Description: The project consisted of bilingual-bicultural programs for both Spanish and Creole French and extended from the first through the ninth grade. Separate classes for both languages did not exist at each grade level. Curricula included were: for the bilingual French classes, reading, math, science, and social studies (Grades 1-6); for the bilingual Spanish classes, phonics, reading, science, social studies, language arts, culture, and math (Grades 1-9); and an outline for English as a Second Language (Grades 1-9).

English language deficiency was determined through testing with the English Language Assessment Battery, developed by the NYC Board of Education. Students were considered language deficient if they scored below the 20th percentile. Spanish-speaking children below the 20th percentile were also given the Spanish Language Assessment Battery. Of 669 children found to be eligible, 212 were placed in special bilingual or intensive ESL classes; the other 457 were served in regular classes with "pull-out" programs which offered ESL instruction. The project consisted of seven bilingual classes in four school buildings. Two were bilingual French-English. Additionally, three schools offered intensive ESL classes (one elementary and two junior high schools). The teachers of all classes received special training from the project staff and ESL materials for language deficient students.

The bilingual classes were initially defined as providing all work in the native language other than for the ESL component; however, some schools' bilingual classes joined the regular classes for art, music, gym, and any other elective courses. As some competence in English was attained, students were moved into mainstream classes, although competence levels were not defined. Additionally, near the end of the school year, the Board became concerned about the transitional aspects of the bilingual program and insisted that the native instruction be restructured to only 50% of the instructional time. Therefore, comparisons of groups and/or methods were questionable, as instructional methods were not held constant.

Pre- and post-testing with SESAT English and the SAT Reading and SAT Math were utilized to compare achievement in the bilingual/ESL settings with achievement in the "pull-out" setting. An analysis of covariance procedure with pre-test scores as covariate was conducted:

1. SESAT English - revealed no significant difference in achievement between the three groups.
2. SAT Reading - revealed no significant differences among the bilingual, ESL, and "pull-out" students.
3. SAT Math - indicated significant differences between the bilingual and "pull-out" groups. The students receiving instruction in their native language achieved higher scores than those whose instruction was in English.

Strengths: The project staff was considered highly qualified and the district was committed to the project.

Discussion: The enthusiasm toward the transitional aspect of moving the students toward mainstream as quickly as possible was not consistently adhered to by the staff. The Board changed its goals near the end of the year and required more English in all settings.

The failure to maintain specified conditions in the treatment as well as the designated control group vitiates any conclusions regarding the preference of one method over another. There were achievement gains for all groups. The proficiency attained in mathematics in the bilingual group was significantly greater than in the "pull-out" program.

While the program evaluators felt that the cultural objectives had been met, there was no documentation that knowledge of the native culture had increased, other than the reviewers' recognition of colorful room decorations depicting culture or language instruction. The curriculum for the bilingual Spanish classes did include a segment on culture.

Reference: Balasubramonian, K., Seelye, H.N. & De Weffer, R.E. Do bilingual education programs inhibit English language achievement: a report on an Illinois experiment. Paper presented at the Seventh Annual Convention, Teachers of English to Speakers of Other Languages. San Juan, Puerto Rico, May 9-13, 1973.

Baker de Kanter Review - p. 42-43, Chapter 2.

Available From: ERIC Document Reproduction Service, ED 118 703.

Name of Study: Do Bilingual Education Programs Inhibit English Language Achievement? A Report on an Illinois Experiment.

Author and Date: Balasubramonian, K., Seelye, H.N. & De Weffer, R.E.

Location: Fifteen schools within three comparable school districts in Illinois.

Treatment Group: 213 Spanish-speaking students (grades kindergarten through three) who received English as a Second Language instruction in the context of a bilingual program.

Control Group: 104 Spanish-speaking students who received English as a Second Language (ESL) instruction within the traditional school program.

Duration: Pre- to post-testing for this study was a five-month period in 1972, January to May.

Ages: Grades K-3.

Type of Program: ESL in transitional bilingual program compared with ESL within traditional school program.

Description: The Balasubramonian et al (1973) study was conducted in response to the fears of parents and educators that the removal of students from the traditional classroom to the bilingual program, resulting in 25% less English exposure and instruction during the school day, might retard English acquisition and educational achievement. The research questions was: Do bilingual education programs inhibit English language achievement?

To probe this concern, pre- and post-test data were collected during a five-month interval in fifteen schools within three comparable school districts in Illinois in 1972. Program alternatives compared were the ESL/bilingual programs and the ESL/traditional approach.

The bilingual program involved the student in a special class placement for one-half of the school day; however, for one-half of that time (or for one-fourth of the school day) the ESL program was given, therefore, only one-fourth of the day was different from that provided Spanish-speaking students in the ESL/traditional classroom. The number of adults within environments did not vary (3); the regular teachers had over twice as many years of experience and the bilingual educators were more fluent in Spanish.

A pre-test, post-test non-equivalent group was used. Data were analyzed by ANCOVA and partial correlational analysis. The authors explain that:

> The present experiment is a situation where the covariate is fallible due to the imperfect reliability of the present measure and there is

> a possibility of a systematic difference on the covariate due to non-random assignment. As a procedure of cross validation, both ANCOVA and partial correlation analysis were performed and Lord's technique of adjustment was analyzed to the 2nd and 3rd grade data, using verbal and non-verbal intelligence scores as covariates. Analysis techniques failed to detect differences.

The results of the listening and reading comprehension tests were compared. Tests used included the following: Test of Basic Experience - Language Subtest (K & L), Test of Reading - Inter-American II, D E, Vocabulary, Speed, Level; Test of General Intelligence - Habilidad General, Verbal, Numerical, Non-Verbal. Pre-test data was obtained in January and post-testing occurred in May.

The analysis revealed no significant difference in the achievement of the two groups. The amount of English achievement in kindergarten and first grade was less than the achievement in first and second grades. Achievement of second and third graders found no significant difference between groups.

The researchers concluded that half-day bilingual programs conducted in this manner do not inhibit English language achievement and the additional benefits of the bilingual program could enhance self-concept and cultural appreciation and awareness, although it was not the purpose of the study to measure these additional benefits.

Strengths: A rather substantiated and appropriate application of analysis techniques occurred.

Discussion: The authors indicated a desire to have used tests reflecting oral language fluency, given the age level considered; however, the constraints made this type of assessment impractical for the project. The use of the selected tests was considered equally suited for the experimental and control groups, since both groups taught were in lower SES Title I schools. Assignment to these two groups was not random; however, the pre-test indicated a high correlation between the two groups. The schools with the bilingual programs were supposedly selected because of the higher percentage of Spanish-speaking children, but it was explained that a lack of funds was the primary reason bilingual programs were only in certain schools. The participants selected for inclusion in the study were students present for both pre- and post-tests. However, there was an attrition factor of 26% project, 21% control group.

The results, while demonstrating the need for further investigation, indicated that placement in the one-half day ESL/bilingual program is not detrimental to the English language development of the participants. Additional rewards for this alternative included cultural appreciation and self-concept enhancement.

Concerns:

1. The interval of testing, approximately four and one-half months, is extremely short for pre-/post-test differences to emerge.

2. Random assignment was not possible since children were already assigned to programs. It is possible that differences that

may be observed in the post-test score are the result of pre-existing group differences rather than treatment effects.

3. The one-half day bilingual instruction as the only treatment may not have been sufficient treatment for effects to be identified.

Reference: Barik, H. C. & Swain, M. Three year evaluation of a large scale early grade French immersion program: the Ottawa study. Language Learning, 1975, 25(1) 1-30.

Baker de Kanter Review - p. 66-67, Chapter 2.

Available From: The Ontario Institute for Studies in Education, Toronto, Canada.

Name of Study: Three Year Evaluation of a Large Scale Early Grade French Immersion Program: The Ottawa Study.

Author and Date: Barik, H. C. & Swain, M.; 1975.

Location: Ottawa, Canada

Treatment Groups: Cohort 1

1970 - ten kindergarten classes with a total enrollment of 219 children

1971 - ten first-grade classes with a total enrollment of 194 children.

1972 - nine second-grade classes with a total enrollment of 170 students

Cohort II

1971 - twenty kindergarten classes with a total enrollment of 422. Data analysis used 50% of the total group.

1972 - eighteen first-grade classes with a total enrollment of 396. Analysis was performed on ten randomly selected students.

Cohort III

1972 - twenty kindergarten classes with a total enrollment of 412. Analysis was performed using ten randomly selected students.

Control Groups: Cohort I

1970 - nine regular kindergarten English instructed classes with a total enrollment of 200 students.

1971 - ten regular first grade classrooms with a total enrollment of 225 students.

1972 - ten regular first-grade classes with a total enrollment of 120 students. Twelve randomly selected students were included in the data analysis.

Cohort II

1971 - twenty-one regular kindergarten classes with a total enrollment of 310. Data analysis was performed on one-half of the students.

1972 - seventeen regular first-grade classrooms. Barik and Swain (1975) do not note the enrollment figures. Ten randomly selected students were used for data analysis.

Cohort III

1972 - thirteen regular kindergarten classes. The authors do not provide student enrollment figures. Of the total students enrolled, data analysis was performed by using ten randomly selected students.

Duration: Three years (from the Fall of 1970 to the Spring of 1973.)

Type of Program: Barik and Swain (1975) employed a French bilingual immersion program. French was used as the only language of instruction in the kindergarten and first-grade immersion classes, in the second-grade classrooms, one hour of English Language Arts per day was taught by an English-speaking teacher. The rest of the instruction continued to be taught in French. The control group had one hour per day of FSL instruction.

Description: The journal article that appeared in Language Learning by Barik and Swain (1975) supplied only a skeletal version of the longitudinal program evaluation of the French immersion school program. The article detailed the number of students in the treatment and control groups, the tests and measures used in the study, and the results of the data analysis. The requirement for brevity in the journal article did not permit the researchers to describe the methods used in teacher selection or qualifications, material selection, or description of the actual program. As a result, it is impossible to gain a full understanding of the program itself.

Barik and Swain (1975) provide a brief description of the communities that participated in the study. Students were from two school districts in Ottawa, Canada. Students in the treatment and control groups were all from middle and upper-middle class English-speaking Canadian homes. Participation in the study was completely voluntary. Children with any visual, hearing, or emotional problems were not included in the study.

Test battery for the three school groups included:

Cohort I

Otis-Lennon Mental Ability Test (Primary 1 Level)
Clymer-Barrett Pre-Reading Battery
Metropolitan Achievement Test
French Comprehension Test

Cohorts II and III

Otis-Lennon Mental Ability Test
Metropolitan Readiness Test
Stanford Early School Achievement Test (Level 1)
French Comprehension Test (1972 edition)

Results

Results of the test batteries for the kindergarten cohorts report no significant differences in numerical or pre-reading skills, or general cognitive development between the French bilingual program classes and the regular English-classes.

Grade 1 - Results on the Metropolitan Achievement Test showed the immersion classes scored significantly lower (at the $p<.001$ level) than the control classes on word knowledge, word discrimination, and reading. No significant difference in mathematics skills was noted for the immersion groups or the control groups.

Grade 2 - Performance of the immersion and control groups on the Metropolitan Achievement Test showed no reliable difference in English skills. In mathematics, the immersion classes scored significantly higher ($p<.05$) on the computation subtests. Scores on the Test de Rendement en Francais and the Test de Rendement en Mathematiques indicate that the French immersion group performed at the fifteenth percentile of the French-standardization group.

Discussion: Some caution regarding the results obtained by Barik and Swain (1975) may be warranted. The researchers reportedly used only ten or twelve randomly selected students for some of the data analysis. The use of Analysis of Variance (ANOVA) and Analysis of Covariance (ANCOVA) techniques on such small numbers of students is questionable.

In summary, the French immersion groups in Ottawa, Canada appeared to be successful for students in kindergarten to Grade 2. The French immersion students were as proficient as the English-speaking control classes in reading and mathematics by Grade 2. In addition, the immersion students seem to have achieved a much greater level of French proficiency than English-speaking students receiving instruction in FSL.

The similarities of the Canadian French immersion programs are in sharp contrast to the characteristics of bilingual Spanish/English programs in the United States. The differences in home language status, socio-economic status, and the educational value of learning a second language severely limit generalizing the results of the Canadian studies to American models. In fact, the researchers of the Canadian study state that one of the most critical differences relates to the "social status" of the native language.

The following illustrates such differences:

1. The treatment and control groups were middle to upper-class English-speaking Canadian students.

2. Although Barik and Swain (1975) evaluated non-equivalent groups of students, the IQ measures showed average to above-average intelligence for both groups.

3. Parental and community support were positive in regard to the establishment of immersion programs in the school.

4. The treatment groups already spoke a high status language - English.

5. Learning French was regarded as positive and academic.

6. Teachers were native French speakers.

In the case of linguistic minority groups, Lambert and Tucker (1972) researched Canadian immersion programs and suggested that instruction alternatives other than an immersion model be provided until reading and writing skills are achieved, or that a completely balanced bilingual program be presented in order that instruction in both languages is conducted.

Reference: Barik, H. C., Swain, M., & Nwanunobi, E. A. English-French bilingual education: the Elgin study through grade 5. Canadian Modern Language Review, 1977, V(33), 459-475.

Baker de Kanter Review - p. 68-70, Chapter 2.

Available From: Canadian Modern Language Review, 1977, V. (33), 459-475.

Name of Study: English-French Bilingual Education: The Elgin Study Through Grade 5.

Author and Date: Barik, H. C., Swain, M., & Nwanunobi, E. A.; 1977.

Location: St. Thomas, Ontario, Canada.

Treatment Group: The treatment group was comprised of English-speaking Canadian students enrolled in a partial French bilingual (PFI) program, St. Thomas, Ontario.

Grade 2 (Cohort IV) 19 students
Grade 3 (Cohort III) 21 students
Grade 4 (Cohort II) 15 students
Grade 5 (Cohort I) 18 students

Comparison Groups: The comparison groups for English and math were comprised of English-speaking Canadians of comparable SES in another school in St. Thomas, Ontario. They received regular instruction.

Grade 2-24 students
Grade 3-18 students
Grade 4-18 students
Grade 5-19 students

Comparison groups in French language skills included indeterminite numbers of students from two locales. Regular program students from

Ottawa, Canada who received French as a Second Language instruction comprised one French comparison group. The second French language comparison group was pupils in Total French Immersion programs in Ottawa and Toronto, Canada.

Duration: Five-year report from a longitudinal study.

Ages: Grade 2 through Grade 5.

Type of Program: A partial French Immersion (PFI) program was employed from first to fifth grade. Instruction was conducted in one language during the morning with instruction given in the other language in the afternoon. In the first and second grade the same teacher provided instruction in French and English language arts. In the third to fifth grade language arts in French and English were taught by different teachers. Instruction in French reading and writing did not begin until the end of the second-grade year. Science was not introduced until third grade. Mathematics, music, French language arts, and science were taught in French. English language arts, physical education, and other non-academic subjects were taught in English.

Description: Test Mastery - The following measures were administered to the treatment and English-speaking control groups in the spring of 1975:

1. The Otis-Lennon Mental Ability Test
2. Canadian Cognitive Abilities Test, Non-Verbal Battery
3. Metropolitan Achievement Test

The treatment and French language control groups were administered the following tests:

1. French Comprehension Test
2. Test de Rendement en Francais
3. Test de Lecture

Discussion: The results from Barik and Swain (1977) do not reflect any predictable trend or pattern. The scores from each cohort fluctuated randomly from year to year in both English and math. It would be logical to assume that the scores from one cohort would reveal a gradual increase or decrease from one year to the next. Instead, the results in Cohort I Grade 3 indicate that the English control group performed significantly better (at the .001 level of confidence) in English than the French bilingual treatment group. No differences were reported in math computation, math concepts, and problem-solving. Since two of the three math subsections involved reading, there is a discrepancy between the scores for reading and the scores for math. In contrast to the significantly poorer English skills demonstrated in Grade 3, Cohort 1 groups performed similarly in English and math for grades 4 and 5.

In Cohort II, the Grade 2 English control group performed significantly better in word discrimination, reading and spelling (at the .01 level of confidence). Scores for Grades 3 and 4 of the same cohort are not significant. No significant differences were reported in math for Grades 2 and 3 of Cohort II. The Grade 4 English control did better in math (at the .001 level of confidence in computation) and .01 level for the total math scores. Math subtests requiring verbal ability (e.g. concepts, verbal items), reported no significant differences.

A more extreme illustration of unexplained random differences was evident for the Cohort III group. In Grades 1 and 2 no significant differences were reported for either math or reading. In Grade 3 the English control group performed significantly better in reading (at the .01, .001, and .05 levels of confidence) on subtests. The same phenomenon occurred in math. The Grade 3 English control group did significantly better on all the math subtests. Differences were reported at the .001 level of confidence in computation, concepts, and total math. Even though Barik and Swain (1977) adjusted the mean scores in math and reading by holding age and IQ constant, the unpredictable performance of the treatment cohorts raises questions.

The first question to be raised is why were the Grade 2 treatment groups tested on instruments that required the ability to read? In the description of the program, Barik and Swain (1977) note that reading instruction was not begun until the end of Grade 2 for all of the treatment groups.

The second question relates to the extent that a curriculum or quality of teaching might have varied from year to year. Significant differences in instruction have introduced a variable that may explain the random extremes in the performance of the groups.

Unfortunately, the researchers did not design or conduct this investigation with either appropriate performance measures or adequate control of extraneous variables. Therefore, the results obtained do not appear to be reliable enough to answer the research question asked. In fact, Barik and Swain (1977) conclude that the inconsistent findings of their investigation warranted "detailed analysis of the curriculum documenting the nature of the materials in use, the teaching strategies and the pedagogical approaches" used.

Reference: Campeau, P.L., Roberts, A.I.H., Bowers, J.E., Austin, M. & Roberts, S.J. The identification and description of exemplary bilingual education programs. Palo Alto, California: American Institute for Research, August, 1975 (Corpus Christi component of report, D 57-76).

Baker de Kanter Review - p. 16-17, Chapter 2.

Available From: ERIC Document Reproduction Service ED 123 893.

Name of Study: The Identification and Description of Exemplary Bilingual Education Programs.

Author and Date: Campeau, Peggy L., Roberts, A.O.H., Bowers, J.E., Austin, M. & Roberts, S.J.; 1970-71 through 1974-75.

Location: Corpus Christi Independent School District, Corpus Christi, Texas.

Treatment Group: 143 students in three elementary schools beginning with six kindergarten classes in 1970-71 and adding one grade level per year until a population of 519 students was being served in 1973-74 comprised the initial treatment group. Data collection was consistent only through the end of the 1973-74 school year when 519 students were enrolled in the project in grades 1-3 in three schools (Crockett, Travis, and Evans). Those enrolled in the treatment group from 1970-1974 are indicated according to grade level below.

Year	Kindergarten	First	Second	Third	Total
1970-71	143				143
1971-72	164	126			290
1972-73	173	155	107		435
1973-74	154	148	124	93	519

Control Group: Three elementary schools served as control schools. The schools were selected by the Bilingual Project Staff due to their close proximity to Project schools; the building principals designated one class at each grade level as a control. The Project description stated that classes forming control were approximately equal to the project group in terms of low SES. Although the report does not specify the exact number of control students, analysis of the tables provided for the various grade level performances on tests indicates that in 1972-73 there were 207 controls to 321 Project students tested on at least one scale.

Attrition was not addressed consistently and the numbers tested varied on each scale. Absenteeism in both groups was extreme, with an average of 353 per class or 12.7 per pupil. This rate of absenteeism possibly contributed to considerable variations in the number of students evaluated on each measure.

Class size of the controls was 23 in comparison to 29 for Project students in the kindergarten level; otherwise, at grades 1 and 2 controls and Project students were equal in terms of class size.

Duration: Test data was collected consistently for 1970-71, 1971-72, 1972-73, and 1973-74 school years. The Project continued during the 1974-75 school year, but the Project staff responsibilities were realigned and the assessment/evaluation process was restructured.

Ages: Grades involved were kindergarten, first, second, and third.

Type of Program: Transitional bilingual education.

Description: Although the term "transitional" is not used, students were moved as quickly as feasible to increase English usage. The type of individualization that took place makes the Project classroom programs difficult to describe and impossible to replicate. A statement from the Project description illustrates the individualization efforts:

> "In kindergarten, instruction in English was initially established at 10% of day; Spanish at 90% of day. Individual differences in children necessitated adjustments in language emphasis and by the end of the kindergarten year many children were receiving instruction in both languages for equal amounts of time. Before the end of grade 1 this is true for all children."

Kindergarten children were not screened for eligibility and participation was voluntary. Initial teacher observation and reliance on Spanish for communication was the basis for grouping and re-grouping of students. There was not a description of the selection process for the control students, other than the information that a nearby school was selected (with equivalent SES) and that the building principal selected one class at each grade level of project operation.

Variation occurred in project classes instructional methods. "In one school classes are self-contained, in another all classes at a grade level are in 'open space' which entailed large grouping and cooperative teaching with daily trading of teachers for subject areas." This type arrangement was accomplished in a circular building with pie-shaped classes and a resource center in the middle of the building. The number of volunteer hours to assist individuals also varied. Involvement increased from almost non-existent to over 3300 hours of help to classrooms (supervise playground, chaperone, individual reading and conversation, provide snacks, tell stories, assist in materials preparation and field trip supervision). Information about control class curriculum, schedule, assistance for aides/volunteers is not included, except for one statement: "Until the 1974-75 academic year, control students were instructed monolingually in English. Heavy emphasis was placed on reading and math."

The Corpus Christi Bilingual Program was selected from one hundred seventy-five programs as one of four exemplary programs by the Dissemination Review Panel of the Education Division (DHEW). The introduction to the report states that "these four programs revealed sound evidence demonstrating significantly improved student outcomes." Since the focus of the report was on programming aspects, there is considerably more information regarding texts utilized, daily schedule, staff training, cost factors, and individualization techniques involving aides and volunteers than on research aspects of the project. In fact, details regarding data analysis are sketchy and the text states that evaluation data has been "summarized".

The objective of the Corpus Christi bilingual project was to provide an effective bilingual-bicultural program for children of Mexican-American descent. The four ways of reaching the goal specified were:

1) Use of instructional strategies and routines selected specifically for the program;

2) Review, selection, adaptation and/or development of appropriate materials;

3) Implementation of a staff development program tailored to program strategies and teacher needs;

4) Direct, intensive involvement of parents through voluntary participation in a variety of program activities that depend on their help.

The major features of the program were as follows:

1) Use of bilingual teachers exclusively at kindergarten and first grade, with pairing of monolingual with bilingual teachers for cooperative teaching in grades 2-4.

2) Experienced bilingual Instructional Consultants provided personal assistance and weekly visits to teachers.

3) Daily cooperative planning among teachers and aides by grade level.

4) Administrative staff conducted planning, coordination, and monitoring.

5) Introduction of English reading readiness and aural/oral skills as soon as child demonstrates minimal aural comprehension of English. There was no indication of how this was measured or criteria of "minimal" comprehension.

6) Emphasis on English reading and communication skills for the duration of the program.

7) Use of bilingual teacher aides for instructional support and reinforcement.

8) Use of teacher-developed curriculum materials tailored to participants' needs.

9) Provision of intensive pre-service training for all project staff, reinforced by monthly in-service training during the school year.

10) Promotion of parent commitment to the program through direct involvement in classroom related activities which included parent education meetings, Parent Advisory Council, classroom involvement, (3300 hours of help in the supervision of children), materials exchanges and tutoring.

Evaluation design: Students were administered pre- and post-tests each year. Means and/or grade equivalents were compared for project and control students.

The following table reflects the scales and when administered:

Tests Administered	Grades Administered 1972-73	Grades Administered 1973-74
Stanford Early School Achievement Test	K	K
Inter-American General Ability Spanish	K	K
Inter-American General Ability English	K	K
SRA Achievement Test Primary I	1	1
SRA Achievement Test Primary II	2	2,3
Inter-American Reading Spanish Level I	1,2	1,2
Inter-American Reading English Level I	1,2	1,2
Inter-American Reading Spanish Level II	2	2
Inter-American Reading English Level II	2	2

To verify teacher opinion regarding initial selection of students who were Spanish-speaking, the Oral English Proficiency Test was administered to a 10% random sample. There was no information regarding the outcome of this administration.

Project staff utilized a basic comparison design in 1972-73 and 1974-75 involving administration of pre-tests and post-tests to both project and control groups within each grade. Evaluators point out that the two groups may not show sampling equivalence, although the narrative states that an "attempt was made to balance both groups as closely as possible in terms of ethnic mix, maturation, and the test instruments administered." Matching was made at the school level by the building principals. Consideration was given to the similarity in SES and grade level. Building principals selected an existing class to serve as a control.

Of the families in the Project, 81% had incomes within the poverty level. The average family had 4.5 children. An extremely high absentee rate was present, particularly at the kindergarten level. In the Project, 92% were Spanish surnamed and, according to information from the parents, 74% were native Spanish speakers, and 50% relied mostly on native Spanish for communication. Teacher observation and assessment of reliance on Spanish for communication was the basis for in-class groupings designed for greater individualization.

The Project program at the kindergarten level consisted of two hours daily of language development and language arts, English and Spanish. The kindergarten teachers in 1972-73 at control schools averaged ten years experience while six kindergarten teachers at Project schools were all first or second year teachers. Emphasis was placed on developing enthusiasm for communicating in English and Spanish in all activities. Teachers did not correct usage in either language or evaluate performance formally. English reading readiness was introduced as kindergarteners developed ability to learn these skills. 80% of the students were provided reading readiness in kindergarten and the other 20% received reading readiness instruction in the first part of the first grade. Children spent 45 minutes a day on Spanish social studies (Spanish culture and heritage) and 45 minutes on mathematics. Phonetic analysis skills were taught first in Spanish.

In other classrooms children were grouped for language dominance and achievement. Children worked in 3-4 in-class groups for each subject with rotation of teachers/aides/volunteers every 20 to 30 minutes. The process was highly individualized with teacher and aide giving special assistance and at least one parent volunteer in each classroom to hear children read aloud, converse, interact, or work independently.

Results

As stated previously, evaluation data was summarized. However, data strongly suggest that the bilingual program results in "superior achievement in both languages by project pupils in comparison to controls" (p.D-62). Results of evaluations are reported by grade level for 1972-73 and 1973-74.

1972-73: Scores are reported by mean pre-test and post-test and post-test adjusted for pre-test mean group differences on the verbal, numerical and non-verbal parts of both Spanish and English versions of the Inter-American General Ability Test.

Kindergarten: Project pupils showed significantly higher adjusted post-test scores on the Inter-American General Ability Spanish Test than did controls. No significant difference was found on two Inter-American English subtests nor on three of the Stanford Early School Achievement Tests.

First Grade: Project pupils showed significantly higher (.01 level) adjusted post-test scores on both subtests of the Spanish and English versions of the Inter-American Reading Tests and also achieved significantly higher (.01 level) on all 3 subtests of the SRA Achievement Tests.

Second Grade: Project pupils showed significantly higher (.01 level) adjusted post-test scores on both subtests of both versions of Inter-American Reading Tests and two of three SRA Achievement Tests, with but one subtest exception (the comprehension part of the Spanish version of the Inter-American Reading Test).

1973-74: The same instruments were utilized as the previous school year but post-test covariance adjustments were not calculated. Program effectiveness was demonstrated by significant mean gain differences favoring project groups.

Kindergarten: Project pupils showed significant gains in comparison to controls on the Spanish version of the Inter-American General Ability Test, but were equivalent to controls on the Stanford Early Achievement Test.

First Grade: Project first graders outperformed controls by approximately 0.5 grade equivalents on the SRA Achievement Tests (sig. levels are not reported) and significantly outgained controls on both versions of the Inter-American Reading Tests. Project pupils averaged significantly above the grade level.

Second Grade: Project pupils outperformed controls with respect to average grade equivalents on the SRA Achievement Series (though mean differences were not statistically significant). Project pupils significantly outgained controls on the Spanish version of the Inter-American Reading Tests. The reverse was true on the English version of the Inter-American Reading Test (controls significantly outgained project pupils).

Third Grade: All differences on the three subtests of the SRA Achievement Test favored the Project group (0.4 grade equivalents). Project pupils significantly outgained the controls on both versions of the Inter-American Reading Tests.

Since the same tests were used both years to measure English and Spanish, longitudinal research was possible. The additive effects of the project were considered by comparison of scores of 213 children remaining in the program with those entering the project at a later grade. Results of the comparison indicated that length of time in the program was significantly related to test performance.

Strengths: The Project involved a large sample size with cohort grouping for successive years. The project utilized individualized approaches.

Perhaps the greatest strengths of the Project were the staff training and parent involvement as advisors, volunteers and supporters.

Discussion: The information cited in the Project description concluded that Project pupils outperformed controls on numerous measures, and the longitudinal analysis indicated that continuation in the program had additive benefits for language development in both languages. Information available regarding the control classes makes comparisons questionable, especially since there were many variables unaccounted for within the Project setting. Some of these included the method of subject selection, instructional methods, pupil/teacher/aide/parent volunteer ratios, and variations in program approaches (self-contained and open rotating classes, cooperative monolingual-bilingual team efforts as well as the all bilingual approach). Additionally, the initial placement criteria and the process of individual matriculation within and between classes based on "minimal English comprehension" is not specified. The amount of Spanish utilized with Project students varied with individual students in their interactive relationships with teacher(s), aide(s), and Spanish-speaking parent volunteers.

Reference: Carsrud, K. & Curtis, J. *Final technical report: ESEA Title VII bilingual project*. Austin: Austin Independent School District, 1980, (Publication No. 79.21).

Baker de Kanter Review - p. 55-56, Chapter 2.

Available From: Austin Independent School District, Office of Research and Evaluation, 6100 Guadalupe, Austin, TX 78752.

Name of Study: 1979-80 ESEA Title VII Bilingual Project

Author and Date: Carsrud, K. & Curtis, J.; 1980.

Location: Austin Independent School District:

Allison Elementary School
Becker Elementary School
Brooke Elementary School
Dawson Elementary School
Govalle Elementary School
Ortega Elementary School
Sanchez Elementary School
Zavala Elementary School

Treatment Group: Project students comprised five language categories: Spanish monolingual, Spanish dominant, bilingual, English dominant, English monolingual. More than five hundred students from kindergarten through sixth grade were involved.

Comparison Group: Non-project students comprised five language categories: Spanish monolingual, Spanish dominant, bilingual, English dominant, English monolingual. Approximately two hundred students from kindergarten through sixth grade were included, with some comparisons being made with the district's overall Mexican American student population. (number unknown)

Duration:	This study focused on the instructional period from September, 1979 through June, 1980, but longitudinal analyses were made over a five-year period, 1975-1980.
Ages:	Grades K through 6.
Type of Program:	Transitional bilingual education.
Description:	This technical report contains the purpose of, procedures for, and results from each instrument employed in data collection relevant to major decision questions, evaluation questions, and student outcome objectives of the 1979-80 ESEA Title VII Bilingual Project. These questions and objectives were:

I What type of bilingual program should the Austin Independent School District have?

What is the nature of the English-Spanish bilingual program (i.e. transitional or maintenance) that the Austin Independent School District currently has?

II Should a maintenance bilingual program such as the Title VII program be funded locally if no federal funds are available?

What is the achievement in English of students who have been in a bilingual program for a varying length of time (1-5 years)?

Do the students in the Austin Independent School District K-5 bilingual program maintain or improve their Spanish reading skills?

What have been the achievement gains over the last four years for the following groups of children?

Project:

Spanish dominant
bilingual
English dominant (Spanish-speaking)
English monolingual

Non-project:

Spanish dominant
bilingual
English dominant (Spanish-speaking)
English monolingual

III What should be the foci for parental involvement project-wide?

What are the effects of parental intervention through involvement with Experience Based Curriculum units on student achievement?

What are the activities of the community representatives in project schools?

Did attendance at Parent Advisory Group meetings improve during 1979-80 when compared with 1978-79 attendance?

Which services and activities of the parental involvement component are most highly valued by parents, and what additional services would be helpful?

IV Should the Austin Independent School District implement the Experience Based Curriculum?

What is the effect of the Experienced Based Curriculum on

reading achievement
conceptual development

V What components of the program appear most effective?

What were the levels of attainment for each objective?

1. Students in kindergarten will demonstrate an increase in Spanish oral language skills.

1.1 Materials identified by the curriculum specialist to complement the Experience Based Curriculum units will be furnished to all teachers. Spanish language materials will be furnished to all teachers instructing in Spanish. English language materials will be furnished to teachers instructing in English.

2. Students in kindergarten will demonstrate an increase in English language skills.

2.1 A member of the project supervisory team will visit each new teacher's classroom an average of at least once per month and experienced teachers' classrooms on an expressed need basis.

3.1 Demonstration lessons and other types of on-campus activities will be planned and/or conducted by a member of the project staff for each school and documented by a log of these sessions.

7. Title VII students in grades 4-5 receiving math instruction in English will demonstrate greater gains in math skills when tested in English than did students in 1978-79.

8. Title VII students in grades 2-5 will demonstrate greater gains in vocabulary skills in English than did the 1978-79 students.

9. Title VII students in grades 1-5 receiving Spanish instruction will demonstrate significant gains in Spanish comprehension and vocabulary.

10. Title VII students in grades 2-5 receiving English reading instruction will demonstrate more significant gains in English reading than did the 1978-79 students.

VI What components of the program appear most effective?

What were the levels of attainment for each objective?

3. Spanish dominant students in kindergarten will demonstrate greater gains in basic concepts (tested in Spanish) than

was true of kindergarteners in 1978-79.

4. English dominant students in kindergarten will demonstrate greater gains in basic concepts (tested in English) than was true of kindergarteners in 1978-79.

5. Spanish dominant students in kindergarten will demonstrate greater acquisition of quantitative concepts when tested in Spanish than was true of kindergarteners in 1978-79.

6. English dominant students in kindergarten will demonstrate greater acquisition of quantitative concepts when tested in English than was true of kindergarteners in 1978.

Although this is an incomplete list of the decision and evaluation questions and objectives, the above is a complete list of those questions and objectives noted and addressed in the Carsrud and Curtis report. The eleven areas of data collection include:

PAL Oral Language Dominance Measure
California Achievement Test
Boehm Test of Basic Concepts
Spanish Reading Test (Prueba de Lectura)
Community Representative Activity Logs
Documentation of Staff Development Activities
Parent Interviews
Supervisor's Record of Classrooms Visited
Documentation of Materials Sent to Schools
Community Advisory Group Sign-In Sheets
Interviews with School Personnel

The authors, Carsrud and Curtis, clearly describe how students were identified as being in one of five language categories, but little information was provided on how students were selected for the project. It appears the only criteria for involvement in the project were: 1) enrollment in one of the schools which sponsored the ESEA Title VII Bilingual Project; 2) parental desire for their children to participate in the project; 3) the students' own Hispanic heritage.

Strengths: It appears one of the strengths of this study is that it included a variety of data collection techniques and sources.

Discussion: Primary Acquisition of Languages

The Primary Acquisition of Languages Oral Language Dominance Measure (PAL) was administered to collect data to determine what components of the bilingual program appeared most effective. Student outcome objectives were listed as: 1) Students in kindergarten will demonstrate an increase in Spanish oral language skills; and 2) students in kindergarten will demonstrate an increase in English language skills.

In the fall, kindergarten students who reported a non-English language to at least one item on the Survey of Home Languages were administered the PAL as a pre-test. In the spring, a sample of project and non-project kindergarten students were administered the PAL as a post-test. Apparently, many students were pre-tested and not post-tested, but the sample comparison of pre- and post-test scores included 88 students tested in English and 78 students tested in Spanish. The sampling procedure was not explained, but it was noted in the report:

> Although this sample included students from both project and non-project classrooms, there were no non-project students in the bilingual or Spanish dominant categories.

In other words, the comparison group consisted of students representing only one language category, English monolingual or dominant, while the experimental group consisted of students representing three language categories: Spanish monolingual or dominant; bilingual; and English monolingual or dominant. ANOVA comparisons between project and non-project English dominant students were made, but comparisons between project and non-project students in other language categories were not made because of the lack of bilingual and Spanish dominant non-project students in these categories. Results indicate that the project and non-project English dominant groups did not differ in their rate of Spanish gains.

When comparing pre- and post-test PAL scores, two groups of students showed significant gains in English: non-project English dominant students and Spanish dominant project students. Bilingual and English dominant project students did not show significant gains. In reaction to the finding that non-project English dominant students improved more than project students, they noted:

> It appears that special attention may need to be paid to the English language acquisition of bilingual and English dominant students in the bilingual program in future years, so that these students do not fall behind in their English verbal ability.

The reliability and validity of this data is uncertain. Although the PAL was administered under standardized conditions in the spring, the researchers noted, "In the fall, conditions are supposedly standardized, but this standardization is not certain." Only five to six months elapsed between fall pre-testing and spring post-testing. Inter-rater correlations ranged between .71 and .94 for five raters in previous years, but researchers indicated that the correlations would be lower for the 1979-80 program year.

California Achievement Test

The California Achievement Test (CAT) was administered to a sample of students in grades 2-5 in order to address the following four aspects of the 1979-80 Title VII Bilingual Programs: 1) Overall student achievement including a comparison of project and non-project students, and gains for the current year; 2) Experience Based Curriculum (EBC), Early versus Delayed Treatment groups; 3) EBC Parent Training; and 4) Impact of the program on long-term academic achievements.

The EBC consists of activity oriented units produced by the project to teach communication skills through introductory and follow-up lessons associated with field trip experiences. The Early Treatment Group was comprised of project students who were exposed to the EBC units and field experiences from September, 1979 to February, 1980, while the Delayed Treatment Group was comprised of project students who were not exposed to EBC units or field experiences until after February, 1980.

Three parent groups were involved in the EBC parent training. The Volunteer Participant Parent Training Group were parents who volunteered and were actually chosen to participate in the parental involvement activities. The Volunteer Non-Participant Parent Training Group were parents who also volunteered to participate in the EBC parental involvement activities, but were placed in the group of randomly selected volunteers who did not participate. The Non-Volunteer Parent Training Group were parents who did not volunteer to participate in any of the EBC parental involvement activities.

Reading subtests were administered to a sample of project students and non-project students in grades 2-5, while math subtests were administered only in grades 4 and 5. The sampling procedure was explained in detail and appears adequate and appropriate. Both project and non-project groups were represented by four language categories: Spanish dominant, bilingual, English dominant (Spanish-speaking), and English monolingual. However, approximately 75% of the project students were either English dominant or English monolingual. Concerning problems with test administration that might have affected the validity of the data, researchers noted, "The school personnel who administered the pre-test in previous Aprils may or may not have followed consistent procedures." This study did not report validity data.

Both project and non-project students in Grades 2-5 showed significant gains in their Reading Total scores from April, 1979 to February, 1980. On Math Total scores, 4th and 5th grade project students also showed significant pre- to post-test gains; 4th grade non-project students did not show significant gains in their math scores. A comparison of gains made by project and non-project students indicated the two groups did not differ at any grade level on the CAT Reading Total. On Math Total, project students gained more than non-project students at 4th grade, but not at 5th grade. Although both groups gained in their raw scores, their scores generally declined with respect to national percentiles.

Comparisons between Early and Delayed EBC Treatment groups on the Vocabulary, Comprehension, and Reading Total Scales of the CAT indicate that the two treatment groups did not differ on any of the scales at grades 2-4; at grade 5, the Delayed Treatment group made greater gains than the Early Treatment group on the Vocabulary and Reading Total scales of the CAT. Researchers noted:

> It appears that the EBC units were not any more effective, and perhaps were less effective, than the curriculum they replace. The results of the fifth grade may reflect an even greater need at this level for complex and sophisticated materials which emphasize achievement related skills.

Comparisons between the three EBC parent training groups indicate there were no differences between the three groups on the Vocabulary, Comprehension, or Reading Total scales of the CAT. However, researchers noted:

> ...the students whose parents volunteered to participate in the parent training scored consistently higher at both pre- and post-testing than students whose parents did not volunteer. It appears that parents do make a difference.

Boehm Test of Basic Concepts

The Boehm Test of Basic Concepts was administered to all kindergarteners in September, 1979, and retested in February, 1980. The purpose of the testing was to provide information addressing overall kindergarten achievement, EBC impact on Early and Delayed Treatment groups, and EBC parental involvement activities impact on academic achievement. The test was administered by classroom teachers who may or may not have received training on the administration of this particular test. Researchers also noted that "Individual variations in administration procedures may have occurred." Students receiving the majority of their instruction in Spanish took the Boehm in Spanish, all others took the Boehm in English. Teachers made the decision as to which language to test in, and pre- and post-testing were done in the same language. Two language categories were represented in the testing: Spanish dominant and English dominant. However, the comparison group consisted of only English dominant students. Results of the Boehm analysis indicate that the project students gained more than the non-project students on all Boehm subscales, except for Time. The results also indicated greater gains than Delayed Treatment groups on their Boehm Total scores, while for groups tested in English, Delayed Treatment students performed better than Early Treatment students. The three parent training groups in the EBC Parental Involvement component did not differ in their gains on Total Score.

Spanish Reading Test

The Spanish Reading Test, or Prueba de Lectura (PAL), was administered to 3rd, 4th, and 5th grade project and non-project students who had taken the test in the Spring of the previous school year. It is unclear how many students were excluded from testing because they had not been tested the previous year. Testing was done in March, 1980, and although some make-up sessions were scheduled, they were not always possible. Three language categories were represented in the testing: Spanish dominant students who received reading instruction primarily or solely in Spanish; English dominant students receiving reading instruction in English only; and English dominant students receiving some reading instruction in Spanish. However, there were no Spanish dominant project students in the 4th and 5th grade samples.

3rd grade English dominant and Spanish dominant students who were receiving Spanish reading instruction made significant gains on their Total Score on the PAL. Non-project students, and those English dominant students who did not receive Spanish reading instruction did not make significant gains in their PAL scores. Regression analyses were conducted to compare gains of project and non-project students. At grades 3 and 4, project students showed greater gains than non-project students. In grade 5, project and non-project students did not differ in their rate of gains.

A longitudinal analysis of third, fourth, and fifth grades was also done with twenty-five to eighty-three students who were consistently project or non-project over a three-year period, 1977-1980. Although project students tended to score higher than non-project students, the authors noted:

> An inadequate sample of consistently non-project students precluded a comparison of project and non-project students... since the sample of non-project students was small, estimates of means for that group may be less accurate than for project students.

Community Representative Activity Logs

Each school's community representative maintained a log which consisted of a checklist detailing all of their daily activities. No mention is made of how these representatives were selected or their qualifications. A random sample of five dates was chosen for examination in the analysis of community representatives' activities. However, some schools were unable to provide data due to prolonged vacancies in community representative positions throughout the year. Analysis of the available data indicated community representatives spent most of their time making home visits. The main purpose of the home visits, telephone calls, and other activities was to improve school attendance, interview parents, and secure clothing, food, services, etc. for the students.

Documentation of Staff Development Activities

The documentation consisted of a listing of staff development activities conducted by Title VII during the school year. There was a total of 199 sessions listed, but one-third of these only involved one participant. Three-fourths of the sessions focused on training in either bilingual education methods or general teaching methods.

Parent Interviews

A random sample of ninety-seven parents was interviewed by community representatives for the purpose of determining which services and activities were most beneficial to parents, and to secure feedback and suggestions concerning the services provided by the community representatives. Data was not available from some schools due to vacancies in community representative positions throughout the year. The authors noted:

> The results of these interviews indicate that the training of parents by community representatives and their use of the home (EBC) units had no measurable or consistent effect upon the parents' knowledge of or attitudes toward school and bilingual education.

The data also verified the activities of the community representatives as noted in their daily activity logs and parents indicated a need for more of the same kind of services.

Supervisors Record of Classrooms Visited

Title VII supervisors kept a record of the date of each visit they made to a project classroom throughout the year. A comparison with recorded visits from previous years indicated an increase in visitation, but it was not possible to determine from the available data whether or not the supervisors' visits actually corresponded to the expressed needs of the teachers visited. The 191 visits only average out to about two teachers visited per school each month.

Documentation of Materials Sent To Schools

The Curriculum Specialist for Title VII maintained a listing of all materials sent to the project schools. The listing reveals that the materials distributed to project teachers were four EBC units which focused on: 1) Pollution; 2) Nutrition and Mexican American foods; 3) The Vaquero; and 4) The Wide World of Sports.

Community Advisory Group Sign-In Sheets

These sign-in sheets contain the name and other information for persons attending each of the eight Title VII Community Advisory Group meetings. Results indicate no significant changes in attendance when compared to previous years, with a current attendance average of forty-one. However, it was noted that everyone present may not have signed in.

School Personnel Interviews

Thirty-one Title VII teachers were randomly selected to be interviewed in April and May, 1980 by persons from the Office of Research and Evaluation. Results indicate an average class size of 22.6 with 1.4 students in each class being Spanish monolingual; 17 of the teachers interviewed had no Spanish monolingual students, while the other fourteen teachers had from one to seven Spanish monolingual students in their classes. Teachers spent an average of 82.5% of the time teaching in English, with ten of the teachers indicating they taught only in English. Only half of the teachers interviewed had bilingual certification. Teachers' perceptions of what the bilingual program was and what it should be were diverse, but they commonly expressed a strong need for instructional materials, coordination between the bilingual curriculum and the AISD curriculum, bilingual specialists, and support services. There was a diversity of opinion among the teachers interviewed concerning more specific areas of the Title VII program.

Summary

Although it was the intention of the federal legislators for ESEA Title VII programs to improve the achievement of students who enter public schools with a Spanish-speaking background, the vast majority of students in the AISD Title VII program - approximately seventy-five percent - were English dominant or English monolingual. The achievement gains noted in the program seem attributable to these English speaking students, rather than the bilingual, Spanish dominant, and Spanish monolingual students. In addition, teachers provided instruction in English 82.5% of the time, making the program one of "immersion" rather than traditional bilingual education. Only one-half of the teachers possessed bilingual certification. Although the major strength of this evaluation was the variety of input data and methods of data collection, the lack of input and controls in various stages of evaluation and incomparable project and non-project groups make it difficult for AISD to make satisfactory decisions regarding the future of their bilingual program. Carsrud and Curtis noted in their overall evaluation summary:

> If any of the AISD Title VII projects are to be considered for continuation after the ending of external funding, than a major re-analysis of the programmatic activities is needed to improve the benefits obtained for the costs required. The

> final evaluation report does not reflect the possible benefits which the program may have provided in areas other than achievement...It is unclear what specific positive effects may have occurred, or whether the effects might have been achieved in some less costly way.

Reference:	Cohen, A.D., A Sociolinguistic Approach to Bilingual Education: Experiments in the American Southwest. Rowley, Massachusetts: Newbury House, 1975. Baker de Kanter Review - p. 51-52, Chapter 2.
Available From:	Newbury House Publishers, Inc. 68 Middle Road, Rowley, Massachusetts 01969
Name of Study:	The Redwood City Project
Author and Date:	Cohen, A. D.; 1975.
Location:	Redwood City, California
Treatment Group:	Pilot group, Fall, 1969 included ten Hispanic male, five Hispanic female, first graders. Follow-up I group, Fall, 1970, included ten Hispanic male and six Hispanic female first graders. Follow-up II group, Fall, 1971, included nine Hispanic male and five Hispanic female first graders. The total number of students included in the three cohorts of the bilingual treatment group was forty-five. The treatment group contained a two-to-one ratio of male to female Hispanic students. Priority consideration for selection in the treatment group was given to students 1) whose family had demonstrated residential stability, 2) whose siblings were already in the program, 3) whose parents gave permission for inclusion in the project, 4) those with the greatest lack of English skills, and 5) children who were Spanish dominant.
Control Group:	The pilot control group contained nine male and five female Hispanic first graders. The follow-up control group included eight male and

seven female first graders. The follow-up II group contained nine male and seven female Hispanic first graders. The total number of children in the control groups was forty-five. The total number of males was twenty-six, females nineteen.

Criteria used for inclusion in the control groups were that the child be: 1) Spanish-dominant, and 2) that Spanish be used as the primary language in the home.

The Hispanic children in the bilingual treatment and control groups were recent immigrants to the United States. Most of the parents were born in Mexico and nearly one-third of the students in the study were born in Mexico.

Duration: 3 years

Ages: Grades K-3

Type of Program: A partial bilingual program model was used. In kindergarten, social studies, music, art, and physical education were taught in Spanish. Spanish dominant students received math and science instruction in Spanish. Language arts was team-taught so that students received instruction in Spanish language arts with one teacher and English language arts with another teacher.

The partial bilingual program model was used in year 2 of the project. Spanish-dominant first graders received math instruction in Spanish. The rest of the treatment groups received math instruction in English. First grade students received science instruction in Spanish. Second grade students were taught science in English. Social studies and music were taught in Spanish.

During the third year, a full bilingual program model replaced the partial bilingual model used in years 1 and 2. Math, social studies, science, and all other subjects except language were taught in Spanish and English in grades 1-3. An alternate days approach was used. Subjects were taught in Spanish one day, English the next. With this approach, a lesson was previewed in Spanish or English, presented in the other language the following day, and reviewed in Spanish or English the next day.

Description: The study followed a quasi-experimental "non-equivalent control group design" (Campbell and Stanley, 1963). In this design, treatment and control groups are not randomly selected and are assumed to be non-equivalent.

Several statistical techniques were employed including: one-way analysis of variance (ANOVA) for measures administered only once; analysis of covariance (ANCOVA) was used to reduce possible bias on post-test comparisons by statistically adjusting pre-test differences; regression analysis was used; and contingency-table analysis was used for categorical data.

Implementation of the program models varied from year to year and from group to group. Cohen (1975) states that the diversity of instruction treatments experienced by the treatment groups illustrates "that bilingual education isn't one thing, but rather a number of possible combinations of treatments." Each of the groups experienced different exposures to bilingual schooling.

Students in the treatment groups did not receive the same bilingual training. For this reason, a description of the instructional program was well-documented for each group and each year of the project. Descriptions of personnel, students, facilities, scheduling, instructional materials, and in-service teacher training overviews were presented. Although the instructional staff over the three years of the projects had different Spanish language backgrounds, a decision was made to speak the variety of Spanish spoken by the children's parents.

The comparison groups of Hispanics received English only instruction. Nearly half of the students in the comparison groups received additional instruction in English as a Second Language (ESL), Title I, or tutorial instruction.

Instruments: Thirteen measures of language proficiency in Spanish and English were administered. Only five of the thirteen measures were given to all groups on a pre-test/post-test basis. Pre-test administrations of five measures occurred in the Fall, 1970 and post-tests were administered in the Spring of 1972. These measures included:

1) Spanish Word Naming by Domain is a test of vocabulary production. It purports to measure a student's ability to name objects commonly found in settings associated with the domains of home, education, religion, and neighborhood. The child is given 45 seconds to name as many Spanish words for objects found in each setting as she/he can.

2) English Word Naming by Domain is an English version of the Spanish Word Naming task.

3) Spanish Storytelling Task is based on the John T. Dailey Language Facility Test (Dailey, 1968). A series of oral stories were elicited based on three pictures--a snap shot of an Anglo woman outside a white house with Chicano, Black and Anglo children clustered around her. A painting by Murillo of "The Holy Family of the Little Bird" was used. In this painting a bearded man is holding a little child, while a woman and a dog look on. In addition, a sketch of a boy pointing at a cat in a tree was shown to the students. The stories were taped and rated by linguistically trained judges. Categories included general fluency, grammar, pronunciation, intonation, language alternation, and descriptive ability. In post-testing a sub-sample of students were presented three photo cards from Words and Action (Shaftel and Shaftel, 1967).

4. English Storytelling Task used the same format as explained above for the Spanish Storytelling Task. Cohen (1975) gives no explanation for using the same stimulus pictures in pre-testing the Spanish and English storytelling task. There is a question of the practice effect of two like tasks administered in the same time frame. The language production of the second pre-test administration of the Dailey stimulus picture could reflect practice effect more than language proficiency.

5. Student Spanish and English Proficiency-Parent Report.

Four reading measures were administered.

1. The Prueba de Lectura Cine-Interamericana was administered to the pilot bilingual group in the Spring of 1971. All groups received

the post-test in the Spring of 1972.

2. The Murphy-Durrell Reading Readiness Analysis was only administered to the follow-up groups I and II in the Fall of 1970 and 1971 respectively. No post-tests were administered.

3. The Inter-American Reading Test was administered to the pilot and follow-up I bilingual treatment groups in the Spring of 1971. All groups were administered the post-test in the Spring of 1972.

4. The Cooperative Primary Reading Test was administered to the pilot and follow-up I groups in the Spring of 1971. Post-tests were administered to the pilot, follow-up I, follow-up II treatment and pilot control groups in the Spring of 1972.

Spanish and English writing samples were collected from the pilot treatment group in the Spring of 1972. As a result of the spotty administrations of pre-test and post-test measures, Cohen (1975) relied most heavily on oral language productions rated by judges as the measure of Spanish and English language proficiency.

Measures for math included:

1. The Mathematics Ability Cooperative Primary Math Test was administered to all groups in the Spring of 1971.

2. The Academic Aptitude Inter-American Nonverbal Ability Subtests were administered to the pilot groups in the Spring of 1971; the follow-up I groups in the Fall of 1970; and to the follow-up II groups in Fall, 1971. The follow-up II groups were administered a post-test in the Spring of 1972.

Attitudes of parents and students were gathered by two instruments. The pre-test and post-test interval was two and one-half years. Parents were given the Language Orientation Questionnaire to obtain parents' reactions to seven reasons for their children to learn Spanish. An English version of the questionnaire was given to obtain parental reactions to seven reasons for their children to learn English. Student attitudes toward English and Spanish were assessed by an item from the Cross-Cultural Attitude Inventory. Student attitudes toward school were measured by attendance records. Students' attitudes were only assessed once after the children had been in the project for two and one-half years.

Spanish-Language Proficiency:

Results from the Spanish Language Proficiency measures indicated few significant differences between the bilingual treatment and English control students. The pilot group of students in the bilingual program did significantly better (at the $p<.05$ level) than the English control group in Spanish reading comprehension, total reading and spelling.

The follow-up I bilingual program students showed more descriptive ability on a storytelling task in Spanish ($p<.05$). A random sub-sample of bilingual program follow-up students received significantly ($p<.05$) higher ratings in grammar on the Storytelling by Domain Task than the English control group.

No significant differences were reported between groups at the follow-up II level.

As some indication of the overall effect of Spanish language instruction, significantly more parents of the bilingual program reported stable or improved Spanish speaking ability of their children. Pilot bilingual program group (p<.05); follow-up I (p<.01); follow-up II (p<.01).

English Language Proficiency

Relatively few significant differences were reported between the bilingual program students and the English control students in English language proficiency. Overall, the English control students in the pilot, follow-up I and follow-up II levels appeared to have a better command of English vocabulary.

The pilot English control group did better in spelling (p<.01) than did the bilingual program pilot students. No significant differences were reported in oral comprehension, storytelling, reading, or other areas in English writing skills.

More significant differences were noted between the follow-up I groups, however. The follow-up I English control group performed better in Word Naming by Domain for "kitchen" (p<.05), total word naming (p<.05), English reading as assessed by the Cooperative Primary Tests (p<.01), and English intonation on Storytelling by Domain (p<.05).

The only significant difference in the follow-up II groups was the English control group's higher rating in descriptive ability on the English Storytelling Task (p<.05).

Mathematics:

No significant differences in mathematics performance were reported between groups for the pilot and follow-up I students.

The follow-up II bilingual treatment group scored significantly better ($p<.01$) than the English control group on the Cooperative Primary Test of Mathematics.

Student and Parent Attitudes:

The children in the bilingual treatment groups viewed the Mexican culture more positively than did the control students (at the .05 level of confidence) for children in their third year of the program. The pilot treatment group, who had been in the bilingual program for three years, also had a significantly better attendance record (.05 level of confidence).

The results of the parent attitude questionnaire showed that the parents of children in the bilingual program were more supportive of their children learning both Spanish and English than parents of the control group children. Parents of children in the bilingual program appeared to believe that the use of Spanish in the classroom would have a positive effect on English language ability. In contrast, parents of the control students seemed to believe that Spanish would not necessarily enhance English learning. Cohen (1975) points out that the differences in parental attitudes may be attributed to the reactions of the Mexican culture and negative language reflected by the comparison parents. The parents and students in the bilingual program appeared to have gained a very positive view of themselves, their culture, and language.

Discussion: Cohen employed a variety of measures. Since many of the measures did not include pre- and post-test information, much of the information obtained is not of a longitudinal nature. In some cases, the results the author obtained seemed unclear. The choice of instrumentation also appears to lack definition. In spite of the instrumentation problems and the methodological questions to be raised in the Cohen study, the overall results are singularly positive.

In respect to his study, Cohen (1975) concludes, "It is still too early to assess the ultimate effects of bilingual schooling in Redwood City. Yet the early indications were that bilingual education in this Mexican American community in California was a viable, significant innovation."

Reference: Cottrell, M. C. *Bilingual education in San Juan County, Utah: a cross-cultural emphasis*. Paper presented at the American Education Research Association Annual Convention. New York, New York, February 4-7, 1971.

Baker de Kanter Review - p. 44-45, Chapter 2

Available From: ERIC ED 047 855, Perry-Castaneda Library, University of Texas at Austin.

Location: San Juan County, Utah.

Treatment Group: 91 students (Navajo, White) from three schools in an experimental bilingual-bicultural program.

Comparison Group: 101 students (Navajo, White) in comparable non-bilingual schools in nearby Anglo community.

Duration: September, 1969 to May, 1970.

Ages: 5-7 years (Grades K-1)

Type of Program: Transitional bilingual education. Navajo (L1) was used as the medium of instruction with extensive use of visual and auditory materials, supplemented by community resources. Students also received ESL instruction. As they developed mastery of English, instruction was given in L2.

Description: The focus of the study was an experimental bilingual-bicultural program in three schools in Utah. Prior to the project, it had been customary for LEP/MEP students in these schools to be retained for two years in the first grade, which was assumed to have a detrimental effect on their academic achievement, self-concept, and their concepts of their own culture. The goals of the program were to:

1. prevent retardation in academic areas: provide instruction in L1 while the students learned L2;
2. build a positive self-image: through a bicultural approach;
3. develop closer communication between parents and teachers; and
4. develop a curriculum which "reflects the needs of people with a rich cultural heritage..."

Hypothesis. The students in the bilingual program (treatment) will perform at or above the achievement levels of (control group) students living in/near Anglo communities.

Kindergarten students were pre-tested using the Anton Brenner Developmental Gestalt Test of School Readiness (BGR) and California Test of Personality (CTP). First-grade students were tested on the CTP and the Metropolitan Readiness Test (MRT). Tests were interpreted for students who had difficulty understanding English. Oral language samples were recorded for English language use and proficiency.

Students were post-tested using a different combination of measures; viz,. kindergarten students were tested on the CTP and MRT, while first-graders were tested on the CTP and Metropolitan Achievement Test (MAT). Oral language samples were recorded again.

The primary method of data analysis was Analysis of Covariance, with the pre-test as covariate. Results indicated that there were no significant differences between treatment and control groups on achievement, but that the group mean for the kindergarten treatment group was higher than the comparable control group. However, statistically significant differences ($p<.05$) were found between the control Navajo and the control Anglo students on the CTP and Metropolitan tests. Other results reported include:

1. The bilingual program was accepted with enthusiasm by parents, and
2. The treatment group students were maintaining a positive self-image, equal to Navajo children who were not integrated into Anglo culture.

Strengths: Since this was an experimental project, there was a greater likelihood that the bilingual program would be implemented with minimal violations of the concept of bilingual instruction. Teachers were trained and their classroom performance observed and evaluated against behavioral objectives developed at Brigham Young University, the site of the staff training component of the study.

Discussion: This study had treatment and "control" groups that were considerably non-equivalent at the beginning of the project. The goal of the program was to decrease these differences, i.e., to bring up the performance level of the treatment group to that of the "control" group. The major advantage of this design appears to be that the results speak for themselves, if factors of internal validity are controlled in the study. For instance, differences in the socio-economic background of the two groups or variations in intelligence would not be a threat, if the program were successful in minimizing differences in achievement of the lower-level treatment group.

Even though the study did not have a specific test of intelligence to measure IQ per se, the Metropolitan Readiness Test data do serve the purpose of controlling for differences in entry-level skills (on the pre-test).

Results of this study demonstrated "no difference" between treatment and control groups, but should not be interpreted in a negative light, quite to the contrary. By bringing up the level of the performance of a lower achieving group of students, the effects of bilingual education may have been favorably demonstrated. Unfortunately, the author does not provide important bits of information that would make the results more conclusive:

1. No statistical data are available on the differences between control and treatment group students at the pre-test level stage. Thus, the extent of equalization cannot be determined at the post-test level.

2. No data was reported on the correlation between the MRT and the MAT and the BGR and the MRT. Since these tests were used as pre- and post-test measures, it is important that they should be highly correlated.

3. Finally, language dominance or language proficiency of students does not appear to have been measured on a test of language proficiency. Consequently, it is impossible to determine the extent of improvement in language development between pre- and post-testing.

In conclusion, it is pointed out that the results of this study cannot be compared with those of other projects without taking into account that the design was significantly different. In this case, "no difference" between the groups on the post-test is a positive outcome of the research, whereas in other projects, the same results are usually indicative of "no effect."

Reference: Covey, D. D. An analytical study of secondary freshman bilingual education and its effect on academic achievement and attitude of Mexican-American students. Unpublished doctoral dissertation, Arizona State University, 1973.

Baker de Kanter Review - p. 4-5, Chapter 2.

Available From:
1) Dissertation Abstracts International, 1973, 4789-A.
2) ERIC ED 079 671.

Location: Phoenix High School, Phoenix, Arizona

Treatment Group: 100 students (Mexican-American) in the ninth-grade bilingual education program at Phoenix High.

Comparison Group: 100 students (Mexican-American) in regular ninth-grade classes at Phoenix High.

Duration: One academic year.

Ages: Ninth grade.

Type of Program: No description of the program was available, except that it was a Title VII bilingual program.

Description: Phoenix High School was granted ESEA Title VII funds for the year 1970-71 to implement an exemplary, secondary bilingual program. This study was initiated to determine whether:

1. Cognitive achievement in academic areas of English, math and reading of ninth-grade Mexican-American students in a bilingual program was significantly different from ninth-grade Mexican-American students in the regular program; and

2. Attitudes toward self, school, peers and teachers were significantly different for the two groups of students.

The null hypotheses stated that there would be no statistically significant differences in either of the two areas mentioned above.

The study used a "true experimental design with equivalent groups," with pre- and post-tests. Two hundred students were randomly selected from a larger group of ninth-graders, and those selected to participate met at least one of the following criteria: limited English proficiency, a bilingual home environment, reading deficiency, and deficiency in English and math. These selected students were then grouped into the treatment and control groups of 100 each. However, the author does not describe procedures used to group students, nor does he describe the tests/methods used to gather data related to the above criteria for selection.

Pre-testing was conducted in early September, 1970, with post-testing in mid-May, 1971. Teachers were given in-service training related to administration of the tests and testing was conducted during class time. Instruments used in the study were: a) Iowa Tests of Educational Development (subtests for correctness and appropriateness of Expression (English) and ability to do quantitative thinking (Math)); b) Stanford Diagnostic Reading Test; and c) Nebraska Attitude Inventory. Tests were selected on the basis of reliability, validity, sensitivity, appropriateness, and objectivity.

The data analysis technique was Analysis of Variance for Unequal n. Results indicated that there were statistically significant differences

($p<.05$) between Mexican-American students in the bilingual program and those in regular classes on English, reading and attitudes toward self, school, peers and teachers, with a higher group mean on the posttests for the treatment group than for the control group. There were no significant differences on math achievement between the two groups. Based on the above findings, the author concludes that Mexican-American students in bilingual programs have significantly higher achievement in English and reading and more positive attitudes than those students in regular education. Their ability to do math is not significantly different.

The study provides a set of recommendations which include the suggestion that the study be replicated with SES as an independent variable, with other populations, with other grades and curricular areas.

Strengths: This is one of the few truly "experimental" designs in that subjects were randomly selected from a larger sample of the ninth-grade population of 379 students who met criteria for inclusion in the study. However, it is not indicated that subjects were randomly assigned to treatment and control groups.

Discussion: While the study appears on the surface to be fairly strong in methodology, a closer examination of the data reveals that some relevant information was not provided, so that it is difficult to agree with certainty with the author's conclusions. First, it was felt that the questions asked in the study were too limited in consideration of the goals of the study (according to the title, to determine the effect of bilingual education on academic achievement and attitudes).

The researcher was interested in establishing significant differences between groups, but results do not offer evidence that such differences were necessarily a consequence of the bilingual program, specifically:

1. The data on attitudes toward self and school did not support the author's conclusion and decision to reject the null hypotheses. Treatment and control groups were significantly different on pre-test measures, so that evidence of the treatment effect is lost by such non-equivalency. Scores did not increase significantly for the treatment group between the pre- and post-tests.

2. There was no discussion of the procedure used to assign students to the treatment and control groups, once the 200 students were randomly selected. Thus, it is not possible to determine if the two groups were, in effect, randomly assigned.

3. There was no information regarding the groups on SES, age, sex, IQ, language dominance or language proficiency.

4. The degree of bilingualism or the method used to determine bilingualism of the home environment was not described. Further, it is not known if, and how much, students were exposed to bilingual education prior to the study.

5. Procedures for test administration and data on test administration were not provided.

6. No information was available on process-related variables such as instructional procedures, teacher competence, curriculum, etc., nor was there any evidence that these were either observed and/or controlled in the data analysis.

Reference: Danoff, M. N. *Evaluation of the impact of ESEA Title VII Spanish/ English bilingual education programs. Volume I: study design and interim findings*. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, 1977.

Danoff, M. N. *Evaluation of the impact of ESEA Title VII Spanish/ English bilingual education programs. Volume II: project descriptions*. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, 1977.

Danoff, M. N. *Evaluation of the impact of ESEA Title VII Spanish/ English bilingual education programs. Volume III: year two impact data*. Palo Alto, California, American Institutes for Research in the Behavioral Sciences, 1978.

Danoff, M. N. *Evaluation of the impact of ESEA Title VII Spanish/ English bilingual education programs. Volume IV: overview of study and findings*. Palo Alto, California: American Institutes for Research in the Behavioral Sciences, 1978.

Baker de Kanter Review - p. 57-60, Chapter 2.

Available From: Perry Castaneda Library, University of Texas at Austin. All documents are on microfiche.

EDC 138 090; EDC 138 091; EDC 154 634; EDC 154 635

Location: Nationwide study of 38 Title VII project sites

Treatment Group: 5311 (according to reviewer's tabulation of reported data) Title VII students (Hispanic, White, Black, American-Indian and American-Asian).

Comparison Group: 2460 non-Title VII students (comparable to treatment group in terms of ethnicity).

Duration: Fall, 1975 to Spring, 1976; in some instances six months or less between pre- and post-testing.

Ages: Grades 2-6.

Type of Program: Maintenance bilingual education (according to project director's perceptions).

Description: The AIR Report was a nationwide study of the impact of the ESEA Title VII bilingual education program. Projects in their fourth or fifth year of implementation were selected for evaluation. The goals of the study included the following:

1. Evaluation of the cognitive and affective impact of bilingual education;
2. Description of the educational process operating in bilingual programs;
3. Identification of educational practices which lead to greater gains in student outcomes; and
4. Determination of per-pupil costs.

Student performance in English and Spanish reading and oral comprehension, mathematics, and student attitudes toward school-related activities were measured on a pre- and post-test design. Methods of data analysis included analysis of covariance, gain score analysis and a comparison of difference in pre- and post-percentile ranks based on the CTBS national scores. Results of the study were reported as follows:

1. There was no consistent significant impact on English language and mathematics scores. Title VII students performed lower in English and at the same level in math as non-Title VII students.

2. Title VII students were at the 20th percentile in English reading and comprehension, and at the 30th percentile in math. The former was lower than the percentile for non-Title VII students, whereas the latter was about the same.

3. No significant differences were found in students' attitudes to school-related activities.

4. No consistent relationships were discovered between teacher characteristics and student outcome.

5. The per-pupil cost of Title VII programs was considerably higher, with an average of $531, than non-Title VII programs with an average of $154 (non-district funds only).

6. The profile of Title VII students revealed that 75% were Hispanic, 7-10% White, 3-8% Black, .2-.5% American-Indian and .2-.5% American-Asian. On the average, students had had 2-3 years experience in bilingual programs prior to the study, and less than one-third of the group were in bilingual education on the basis of need for English instruction; i.e., the rest participated for other reasons, including parental desire for their child to receive bilingual instruction.

Strengths: The study attempts to control for several intervening variables such as teacher characteristics, classroom and project variables and student background including the following:

Teacher characteristics: Teacher credentials, number of years teaching, number of years in bilingual classroom, training in bilingual education, language used at home, language proficiency, and college degree.

Classroom and project variables: Number of project directors and support personnel, language proficiency of personnel, ethnicity, perception of school board attitudes toward bilingual program, project size, and type of bilingual program.

Student background: Ethnicity, language dominance, sex, socio-economic status, age, number of days absent, (hours of instruction received), and number of years in bilingual program.

Large sample size. This was the first nationwide study of the bilingual education program.

Discussion: This study has received considerable criticism in the literature for significant weaknesses in its methodology and data collection which make the results difficult to interpret. In their attempt to evaluate ESEA Title VII projects collectively as a program, the investigators combined data from 38 projects to develop an aggregated profile; in so doing, it was almost inevitable that the significant differences in effectiveness were no longer apparent. Other comments are presented below:

1. Treatment and control sites were not always comparable. For several project sites, it was not possible to obtain control

sites nearby, and those selected were in other parts of the state/nation.

2. Teacher judgment and perception were the only bases for determining language dominance, and no measures of language proficiency were required.

3. The time period between pre- and post-tests was six months in some instances; hardly adequate time for any progress to be detected in student outcomes.

4. Several teachers in the Title VII classrooms did not report language proficiency in both languages. Moreover, teachers' language proficiency and bilingualism do not seem to have been measured by a specific test, but rather by means of a questionnaire.

5. Differences in program characteristics do not appear to be controlled for in the methodology.

6. The study does not take such inter-program variability into account when determining program impact.

The major difficulty with a study such as this is that it is an attempt to evaluate the impact of bilingual education, when in fact the programs being evaluated do not always reflect program characteristics and teacher behaviors that are consistent with the concept of bilingual education. Moreover, in the case of the AIR Report, such attempts are further clouded by methodological constraints that make the data rather difficult, if at all possible to interpret.

Bibliography: Center for Applied Linguistics. Response to AIR Study "Evaluation of the impact of ESEA Title VII Spanish/English bilingual education programs". Arlington, Virginia: Center for for Applied Linguistics, 1977.

Intercultural Development Research Association. The AIR Evaluation of the impact of ESEA Title VII Spanish/English bilingual education programs: An IDRA Response with a summary by Dr. Jose Cardenas. San Antonio, Texas: IDRA, 1977.

Reference: Huzar, Helen. *The effects of an English-Spanish primary-grade reading program on second and third grade students*. Unpublished thesis, Rutgers University, May, 1973.

Baker de Kanter Review - p. 46-47, Chapter 2.

Available From: Rutgers University, New Brunswick, New Jersey.

Name of Study: The Effects of an English-Spanish Primary-Grade Reading Program on Second and Third Grade Students.

Author and Date: Huzar, Helen; 1973.

Location: Perth Amboy, New Jersey.

Treatment Group: 41 Spanish-speaking second-grade students
43 Spanish-speaking third-grade students
84 total (two second-grade classes and two third-grade classes)

Control/
Comparison Group: 40 Spanish-speaking second-grade students
36 Spanish-speaking third-grade students
76 total (two second-grade classes and two third-grade classes)

Duration: One year.

Ages: Second- and third-grade students

Type of Program: Treatment Groups - (1) Second-grade bilingual teacher gave reading instruction in Spanish for 45 minutes and a monolingual teacher gave reading instruction for 45 minutes in English each day. (2) Third-grade classes received the same treatment as outlined in No. 1. (3) Second-grade - two bilingual teachers in classroom with one giving reading instruction in Spanish for 45 minutes and one giving reading instruction in English for 45 minutes each day. (4) Third grade received the same treatment as outlined in No. 3.

Description: Huzar (1973) selected both treatment and control group populations from the same public elementary school in Perth Amboy, New Jersey. Reading instruction in the treatment groups consisted of Spanish and English reading instruction for 45 minutes each day. The treatment group varied as to the speaking ability of the teachers in the classroom. Control group received 45 minutes of reading instruction in English each day. The treatment and control groups used different readers (treatment group used 1970 edition of the Miami Linguistic Readers, while the control used the 1968 edition of the Scott-Foresman Open Highway Series.)

The measures used in this study included: Metropolitan Readiness Test (previously administered), Lorge-Thorndike Intelligence Test (previously administered to third-grade students only), and the Test of Reading: Inter-American Series (administered by the author).

Data gathered in this study were analyzed via "t" tests. Arithmetic means between groups were compared for significant differences.

The results of the study indicate:

1. No significant differences between means of reading scores of second- and third-grade students in treatment and control groups;

2. Boys in treatment groups had significantly different (higher) reading scores than did boys in control groups; no difference for girls was noted;

3. No significant differences in second-grade groups in relation to teacher speaking ability;

4. Third-grade classes having two bilingual teachers had significantly different (higher) reading scores than did classrooms with one bilingual teacher and one monolingual teacher.

Strengths: Huzar used treatment and control groups in her experimental design. She compared groups in several ways, e.g., differences in reading scores, differences in reading scores by sex, and differences in reading scores by teacher speaking ability.

Discussion: Huzar's study is open to several threats. Problems with the design and implementation of the study are itemized here in terms of categories:

1. Sample - The author states that the sample population "came from the Spanish-speaking segment of the population in a section of the city which was classified as disadvantaged. The majority of the parents of these subjects were blue-collar workers, many of whom were on welfare." Controls for serious deprivations and differences among students in groups were not made. IQ was not tested among second-grade students and IQ scores for third-grade students were reported only in terms of means per treatment and control groups (92.1 and 90.0 respectively).

2. Limited Statistical Analyses - The use of means and "t" tests (with one Scheffe contrast) composed the entirety of the data analysis techniques. These techniques are not satisfactory to

determine the true relationship of the data. Central tendencies only indicate trends in the data.

3. Generalizability - This study is limited in terms of its findings being able to be generalized to other settings in other locations because of the single sample within one school.

4. Design - In reading the words of the author concerning the design of the study in terms of the treatment, it is unclear what duration of treatment (reading instruction) the different classes were to receive. It appears that the treatment groups with one bilingual and one monolingual teacher are receiving 90 minutes of reading instruction rather than the 45 minutes in the other groups.

5. Detail - Similarities and differences in the backgrounds of the sample population are not made clear. All of the students are reported to have shared the first-grade experience at the same school, but no mention of controlling for this seems to have occurred. Attrition is not mentioned in this study. Since there are no controls used in this study, the phenomenon of attrition could severely change the characteristics of the sample population, especially since they are considered to be disadvantaged.

6. As stated by the author, "the suitability of the Metropolitan Readiness Test and the Lorge-Thorndike Intelligence Test for bilingual children was questionable.

7. Differences in textbooks and reading instruction methods for the treatment and control groups could have seriously affected

the results.

All considered, the lack of significant differences between groups tends to reflect no superiority of either approach used in this study for bilingual disadvantaged students.

Reference; Kaufman, M. Will instruction in reading Spanish affect ability in reading English? *Journal of Reading*, 1968, *17*, 521-527.

Baker de Kanter Review - p. 18-19, Chapter 2.

Available From: Perry-Castaneda Library, University of Texas at Austin

Name of Study: *Will Instruction in Reading Spanish Affect Ability in Reading English?*

Author and Date: Kaufman; 1968.

Location: New York, New York

Treatment group: 75 seventh-grade Spanish-speaking students

Comparison/ Control Group: 64 seventh-grade Spanish-speaking students

Duration: Study at School A covered school years from September, 1963 to March, 1965 (18 months)

Study at School B covered school year 1963-64 (9 months)

Ages: Seventh-graders

Type of Program: Spanish reading instruction for 45 minutes (4 times per week at School A and 3 times per week at School B)

Description: Kaufman (1968) selected seventh-grade Spanish-speaking students whose average reading grades on the *Metropolitan Achievement Test* administered during their sixth-grade experience were one to two-and-one-half grade levels below the sixth-grade level. In addition, subjects came from homes in which Spanish was spoken.

Treatment in this study consisted of instruction in Spanish with emphasis on specific reading skills in Spanish. The treatment in School A varied from that of School B in frequency of instruction.

The results of the study indicated:

1. School B subjects showed slight evidence of transfer from Spanish reading instruction to English usage.
2. The treatment group acquired greater reading ability in Spanish than did the control group.

Strengths: Kaufman utilized a research design which emphasized treatment and control groups. Statistical measures of ANCOVA allowed adjustment of means of pre- and post-test measures.

Discussion: This study utilized Spanish-speaking students who were "retarded" in their English language ability. Students were assumed to be retarded on the basis of reading scores on the *Metropolitan Achievement Test*. No statement as to the intellectual functioning of the students was made. It was used as a covariate in the statistical analysis, however.

In looking at the treatment and control group sizes at both School A and School B, it is noted that at School A, 30% more subjects were in the treatment group than in the control group. The investigator states that subjects were randomly assigned to these groups. The randomness of this assignment is questionable with such disparity in groups, however. The source of the random distribution bears some explanation. It is also noted that the results from School A proved to be insignificant. The groups at School B were equal and the results tended to show some transfer from Spanish reading instruction to English reading ability. The investigator states, however, that the subjects in the treatment group at School B had a greater degree of proficiency in reading Spanish at the onset and had greater time available for applying acquired skills in reading Spanish than did the control group at School B. The reason for this occurrence is not clear. All of these factors tend to indicate, however, that multiple reasons could be responsible for findings both at School A and School B.

In addition to the above, no mention is made of the quality of instruction at either school. Since this variable constituted the treatment of the study, analysis and controls would seem to have been in order to obtain results which would fairly indicate causative relationships.

Time interval between pre- and post-testing at School A was 18 months and 9 months at School B. Achievement testing was done in October, 1963 at School A and in December, 1963 at School B. Since this testing was done almost in the middle of the school year at School B, one would expect achievement scores to be higher.

Three series of retesting occurred at School A, two series of retesting occurred at School B. Retesting was done with the Durrell-Sullivan Reading Achievement Test and the Cooperative Inter-American Test. The frequency of this testing, particularly at School B, could have seriously affected results. No mention was made of how this retesting was conducted. It seems safe to conclude that the students involved were engaged in "large doses" of retesting. The reason for the retest series is unclear, and may have acted either to produce the effect found by teaching the test, or masking the effect of treatment by adjusting the means of the post-test measures used.

Reference: Lambert, W.E. & Tucker, G.R. Bilingual education of children: the St. Lambert experiment. Rowley, Massachusetts: Newbury House, 1972.

Baker de Kanter Review - p. 62-65, Chapter 2.

Available from: Newbury House Publishers, Inc., 68 Middle Road, Rowley, Massachusetts 01969.

Name of Study: The St. Lambert Experiment

Author and Date: Lambert, W.E., & Tucker, G.R., 1972.

Location: St. Lambert, Quebec, Canada

Treatment Group: There is a discrepancy in the number of children reported to be in the pilot bilingual treatment class. Table 1 (page 12) reports a treatment group n of 26 in September 1966. Table 2 (page 13) reports a treatment group n of 22 in September 1966. The n of the pilot treatment group in Grades 2,3, and 4 is not reported. Attrition figures for the pilot treatment group were not reported.

Thirty-eight students are reported in the follow-up bilingual treatment group, Grade 1, 1967 (Table 8, page 48), an n of 36 is reported in Table 7 (page 47), and Grade 1, Table 9 (page 50) reports an n of 34 for the bilingual treatment follow-up class. No subsequent figures are reported for the number of students in the follow-up treatment groups in Grades 2 and 3.

Comparison Group: There is a discrepancy in Tables 1 and 2 of the number of children in the three pilot class control groups. The n for the Canadian English-speaking Control group I is reported as n = 22 in Table 1 and n = 19 in Table 2. Similarly, the n for the Canadian English-speaking Control 11 group is reported as n = 26 in Table 1, n = 24 in Table 2. The discrepancy in the reported ns is not explained and both Tables 1 and 2 report on the same measure, the Raven's Progressive Matrices Test administered in September, 1966.

The Canadian English Control I group n = 21, the Canadian English Control II Group n = 27, and the French Control group n = 21. Table 8 follow-up class comparison on Measure of Intelligence and Components of Socioeconomic Status, Grade 1 report the following: English Control I group n = 26; English Control II Group n = 28; French Control n = 25 (page 43). In the third table, Table 9, Parent Questionnaire for the follow-up classes, Grade 1, the authors report: English Controls, n = 52; French Controls, n = 22. Lambert and Tucker (1972) do not discuss the apparent discrepancy of ns in their study. The authors do report that attrition among the English I and II at the Grade II level groups force them to combine the two classes as one group for statistical purposes. The authors report that attrition among the French Control is so high that city or national norms were used for achievement in mathematics in Grades 2,3, and 4.

Type of Program: A Bilingual French immersion model was employed for the Pilot and Follow-up treatment groups in kindergarten through Grade 1. From the second to fourth grades a partial bilingual program model was used. A French language curriculum was used with two five-minute periods on English Language Arts each day for the bilingual treatment groups. All textbooks, workbooks, and readers were written in French for native French speakers. English was used for instruction in music, art, physical education, and library periods so that nearly 40% of each day's instruction was conducted in English.

Description: The authors contend that treatment and control groups were carefully matched according to a non-verbal I.Q. measure (*Raven's Progressive Matrices*), socio-economic status, home environment characteristics (e.g., emphasis of education, enrichment of home environment, etc.) However, the authors do not describe the procedures they used to select or match the students.

Since it was stated that groups were carefully matched, several points need to be considered: 1) The "matched" control groups in both the Pilot group and the Follow-up group contained different ns. Ordinarily, in matched samples the ns for the treatment and control group would be equal. 2) The authors report that several members of the treatment group had perceptual or other learning deficits. Children with these unusual characteristics would be difficult to match. 3) The statistical treatment, the analysis of covariance (ANCOVA) used in this study would not have been necessary in a matched sample study. These considerations may suggest that while the authors may have matched on the basis of similar means between groups, the groups were not individually matched. Therefore, the groups could have been essentially non-equivalent.

The statistical procedures used by Lambert and Tucker (1972) were analysis of variance (ANOVA) and analysis of covariance (ANCOVA). The ANCOVA procedure used I.Q. and home environment characteristics as covariates. The adjusted mean scores were then tested by analysis of variance.

The authors claim that no attempt was made to document teacher competence, or adherence to the bilingual models used in the study. Lambert and Tucker (1972) also note that no attempt was made to obtain the best teachers. Observations made of all bilingual classes in 1971 by an experienced teacher, Mme. Benorite Noble, seem to indicate the contrary. She notes, "they like, are even enthusiastic about, their work: they are competent, experienced elementary school teachers" (page 242). Mme. Noble's descriptions of the bilingual classrooms are all positive and may suggest that while teaching ability was not used as a criterion for teacher selection, the teachers in the bilingual program appeared to be above average.

Duration: Five years (1966-1971)

Ages: Grades K-4

Procedures:

The pre-tests for the first-grade Pilot and Follow-up groups were administered in the Fall. All subsequent measures were administered in the Spring of each year.

The measures included:

Grade 1--Pilot groups pre-test measures: Fall, 1966.

1. Raven's Progressive Matrices Test.
2. Parent interviews to determine: a) emphasis placed on education, b) quality of the child's linguistic environment, c) guidance in school work, d) enrichment of home environment, and e) educational facilities.

3. Parent Questionnaire on parental attitudes towards French Canadians and the French language. Other questions included years in residence and English and French language proficiency self-reports.

Grade 2--Pilot groups post-test measures: Spring, 1967.

1. Metropolitan Achievement Tests (1959), Primary I Battery.

2. The Peabody Picture Vocabulary Test in English.

3. Word Association Analysis, English and French. Two lists, one in French, the other in English, contained fifty-one stimulus words. Children were asked to give a word that had a similar meaning. All students were administered the French and English word lists. Half of the students were given the English word list first, the other half received the French word list first. After two weeks, the students were given the alternate word list. Each list was coded by two judges who were monolingual in either French or English. The coding was then reviewed and checked by two bilinguals. The authors did not report inter-rater reliability data.

4. Speaking Skill in English. A filmstrip story was presented, the child was asked to re-tell the story. Ratings were made from recordings of expressive ability, grammatical errors, enunciation, rhythm and intonation, time on production, and number of words by two linguistic judges working independently.

5. Speaking Skills in French. A story was presented in French. The child's re-telling was tape-recorded and rated by one French

language teacher. The categories were judged the same as for Speaking Skill in English.

6. Phoneme Production in French. A series of nineteen French phrases, recorded by a French-speaking adult were repeated by the child. The responses were tape-recorded, analyzed, and scored by one judge according to native-like control.

7. Test de Rendement en Calcul--School Commission Mathematics Test in French.

8. Test de Rendement en Francais--Reading Skills in French: Word Discrimination and Sentence Comprehension.

9. Word Discrimination in French.

10. Phoneme Discrimination in Russian. Auditory discrimination task using fifty-three pairs of Russian phonemes.

11. Raven's Progressive Matrices (Sets A, Ab, and B).

Grade 1--Follow-up--Essentially, the same battery of tests used in Grade 1 Pilot Class groups were administered to the Grade 2 groups. The following changes were noted:

1. A French translation of the Peabody Picture Vocabulary Test.

2. Listening comprehension in English was tested by the oral representation of two stories in English to an entire class. The children were asked a total of twenty-four true/false questions about the stories.

3. The word association analysis, French and English. Coding system was defined more clearly. However, inter-rater reliability was still not mentioned.

4. In speaking skills: Story re-telling; word counting; number of adjectives; number of different adjectives; the number of nouns, the number of verbs; the number of different verbs; grammatical errors; overall comprehension of the story theme; and sequence and native-like command of the language rated on a five-point scale by two independent bilingual judges.

5. Speaking skills: Story creation. Children's language was rated and counted in the same way as Speaking Skills, Story Re-telling.

6. Listening comprehension: A French version of test 2 described above.

7. Speaking skills: Story Re-telling in French. A French version of measure 4 listed above.

8. Speaking skills: Story Creation in French. A French version of measure 5.

9. Lorge-Thorndike Intelligence Test, Level 1.

The tests administered to the Grade II Pilot and Follow-up Classes were the same as those used with the first-graders.

Modifications were made in some measures to make them more age appropriate. The measures used with the Grade III Pilot and Follow-up groups were more advanced versions of the same measures. Measures used with the Pilot Grade III were also similar to tests used throughout the study. The following changes were noted:

1. More difficult items were administered of the Peabody Picture Vocabulary Test, French and English.
2. 1970 version of the Test de Rendement en Francais (C.E.C.M.).
3. 1970 version of the Test de Rendement en Calcul (C.E.C.M.).
4. Sets B and C of the Raven's Progressive Matrices Test (1958).
5. Level 3, Form A, of the Primary Battery of the Lorge-Thorndike Intelligence Test.

Results Related to Research Questions:

1. What effect does such an educational program have on the experimental children's progress in home language skills compared with the English controls?

   Grade 1--the bilingual treatment Pilot and Follow-up groups performed poorer (.01 level of confidence) in English reading skills, word knowledge, and word discrimination.

   Grade 2--The bilingual treatment Pilot class performed as well as the English Control group. Only the spelling subtest score was poorer (.05 level of confidence) than the scores made by the English Control groups. The bilingual treatment Follow-up group did as well as the English Controls on subtests of words, knowledge word discrimination, reading, and reading comprehension, and received better reliability scores (.05 level of confidence) on the English version of the Peabody Vocabulary Test.

Grade 3--The reading ability of the bilingual treatment Pilot and Follow-up classes is comparable to that of the Controls.

Grade 4--The reading ability of the bilingual treatment Pilot class was at the same level as Controls. No data was available for Grade 4 Follow-up class.

2. How well do children progress in developing second language skills compared to French children receiving conventional schooling?

   From Grade 2 on, listening comprehension in French for the treatment groups was comparable to the French Control class. Scores by the treatment groups on the French Peabody Picture Vocabulary Test were similar to those of the French Control at Grade 4.

3. How well do children in bilingual treatment groups perform in comparison to Controls in a non-language area such as mathematics?

   The Pilot and Follow-up Treatment groups performed as well in computation and problem-solving as the English-speaking control classes. During Grade 2 the treatment groups did significantly better (Pilot .01 level, Follow-up .05) in computation.

4. What effect does a bilingual program have on the measured intelligence of the treatment classes?

Grade 1--Pilot class scored significantly lower (.01 level) than did the English Control I, Grade 1 class. Results of the Grade 1

Follow-up treatment and Control class showed similar scores on the I.Q. measures.

Grade 2--Intelligence measures are similar for the control and treatment classes. The verbal subtest of Lorge-Thorndike showed the Pilot group to do less well (.05) than Controls. The Follow-up class, however, scored higher (at the .05 level of significance) on the Raven's Progressive Matrices and a subtest of the Lorge-Thorndike.

Grade 3--To determine the cognitive development of the treatment groups, several creativity measures were added to retests of the Lorge-Thorndike and Raven's Progressive Matrices. Creativity measures showed similar results for the Pilot and Controls classes, although the Pilot class performed significantly better (.05) on unusual uses of English.

Intelligence and creativity comparisons for the Follow-up and Control groups revealed no significant differences on either the Raven's Progressive Matrices or the Lorge-Thorndike total scores. The Follow-up group did perform better (at the .05 level) on one of the subtests of the Lorge-Thorndike.

Grade 4--Intelligence and creativity measure for the Pilot class were generally comparable to control groups in Grade 4. There were two exceptions: 1) the pilot class scored significantly better (at the .05 level of confidence) on the mathematics subtests of the Lorge-Thorndike, 2) the English control groups performed significantly better (at the .05 level of confidence) on the Raven's Progressive Matrices Test. Thus, at the end of Grade 4 the effects of bilingual instruction appeared to have no negative effect on either the intelligence or creativity the participants.

In summary, the bilingual treatment groups (Pilot treatment and the Follow-up treatment) appeared to be as proficient as the English control classes on measures of creativity, intelligence, mathematics, and English language skills (L.I.). In addition, the bilingual treatment groups seemed to have developed a remarkable competence in second language French written and oral expressive skills.

Strengths: The St. Lambert Study is generally regarded as one of the better examples of an immersion bilingual program evaluation for two reasons: 1) Lambert and Tucker (1972) matched treatment and control groups on two critical variables: I.Q. and socio-economic status. 2) As an additional assurance of equivalent groups, analysis of covariance (ANCOVA) was used to adjust for group I.Q. differences that might bias the treatment effect. As was discussed earlier, ANCOVA may have been necessary for non-equivalence between the "matched" groups.

Discussion: Canadian English-speaking children immersed in French academic instruction became as proficient in English academic areas as the English-speaking children receiving instruction only in English. At the same time, the students in the French bilingual treatment groups in the St. Lambert study seem to suggest the efficacy of bilingual immersion models. However, the authors caution against overgeneralization of the program results to other groups or sociopolitical settings. For instance, the English-speaking children in the bilingual program already spoke a socially perceived prestigious first language, English. The St. Lambert study was initiated at the request of middle-class parents who regarded French/English bilingualism to be socially and politically beneficial to their children. Bilingualism, in this context, was regarded as enrichment, rather than a compensation. In addition, the bilingual program in St. Lambert was not mandated or required.

Several noteworthy differences exist between the French bilingual programs in Canada reported by Lambert and Tucker (1972) and the Spanish bilingual programs in the United States.

1. Classroom observations by Mme. Noble suggest that the bilingual teachers may have been unusually good teachers. The working conditions, attractive middle-class suburban schools, may well have attracted better teachers.

2. All subjects in the bilingual treatment groups volunteered for the program. Parental support was extremely strong.

3. Lambert and Tucker (1972, page 216) state that the model used in the St. Lambert Experiment "is not proposed as a universal solution for all communities or nations planning programs of bilingual education." The authors go on to state that, "If A is the most prestigious language, then native speakers of A would start their schooling in language B and after functional bilingualism is attained, continue their schooling in both languages." The program model thus advocated by Lambert and Tucker is essentially the one proposed for Spanish bilingual children in the United States. That is, Spanish competency in reading, writing, and oral

production should be attained first with instruction in English occurring either in conjunction with or subsequent to Spanish instruction.

Reference: Legarrata, D. The effects of program models on language acquisition by Spanish-speaking children. *TESOL Quarterly*, 1979, *13*(4), 521-534.

Baker de Kanter Review - p. 13-15, Chapter 2.

Available From: Perry-Castaneda Library, University of Texas at Austin

Name of Study: *The Effects of Program Models on Language Acquisition by Spanish-speaking Children*.

Author and Date: Legarrata; 1979.

Location: Large West Coast city (not named).

Treatment Group: 52 five year-old kindergarten students were involved in the five groups comprising this study.

Comparison/ Control Group: Statistical measures used strongly relied on students in Group 1 which were essentially "untreated" receiving English instruction with no ESL.

Duration: Six months

Ages: Five-year-old kindergarten students.

Type of Program: The effects of five program models were studied:

Group 1 - Traditional Method (English only with no ESL)

Group 2 - Traditional Method with ESL (English only)

Group 3A - Bilingual Method (concurrent translation comprised of 72% English and 28% Spanish instruction)

Group 3B - Bilingual Method (alternate immersion comprised of 50% English and 50% Spanish instruction)

Group 4 - Bilingual Method with ESL (concurrent translation of 72% English and 28% Spanish instruction)

Description: Legarrata (1979) found that significant differences occurred among concurrent bilingual approaches and alternate immersion bilingual approaches. This understanding caused the researcher to divide the original Group 3 into two parts representing the differences in Spanish-English usage.

Four 11- and 12-year old girls gathered the data from subjects. Testing was conducted in a manner designed to make subjects feel at ease. The investigator does not clarify if subjects were Mexican-American or of Mexican descent.

Data was analyzed via full score comparisons and analysis of covariance. Correlation coefficients were also derived between pre- and post-test measures. Results from this study indicated:

1. ESL is not effective in enhancing complex communicative competence skills.

2. The highest gains in English communicative competence, Spanish communicative competence and oral comprehension of English were demonstrated by the bilingual method utilizing alternate immersion approach with no ESL (Group 3B)

3. ESL produced greater gains in Spanish used in the home environment.

4. Acquisition of Spanish and maintenance of Spanish is best facilitated by the bilingual method utilizing alternate approach

without ESL (Group 3B)

Strengths: Legarrata used sound multivariate statistical analyses of gain score comparisons and ANCOVA.

Discussion: Although the author utilized appropriate statistical measurements, the samples within each group were very small. For example, Group 3B, which was found to be the most effective program type, was comprised of only seven students. No mention was made in the study concerning the intellectual functioning of the students involved in the study. In addition, no mention was made of controls for teacher quality or whether the students were of Mexican or Mexican-American descent. All of these factors could certainly affect Legarrata's study if not controlled.

In addition to the above, serious questions arise from the investigator's use of four 11- and 12-year-old girls, two of whom are the investigator's youngest daughters, as data-collectors for this study. No mention is made of tests for inter-rater reliability. Such tests would appear to be necessary.

Legarrata took much care in attempting to establish a testing environment which would put the young subjects at ease. However, provision of cookies, peanuts and raisins "throughout the testing" could have seriously affected results of timed tests employed in the study.

Legarrata concludes that the models suggested in this study should be tested further. Replication of this study, eliminating factors which could detrimentally affect results, appears called for prior to conclusions of the efficacy of the models.

Reference:	Lum, J. B. An effectiveness study of English as a second language. Unpublished doctoral dissertation, University of California at Berkeley, 1971. Baker de Kanter Review - p. 34-36, Chapter 2.
Available From:	ERIC ED 070 321
Name of Study:	An Effectiveness Study of English as a Second Language.
Author and Date:	Lum, J.B., 1971.
Location:	San Francisco Unified School District, San Francisco, California.
Treatment Group:	This study contrasted the effectiveness of English as a second language and bilingual education instructional methods in improving the English language maturity of Chinese monolingual/monocultural students. The treatment group was comprised of thirty-five monolingual Chinese students who received bilingual instruction.
Comparison/ Control Group:	There were twenty-five monolingual Chinese students in the control group which received ESL instruction. Monolingual/monocultural students were selected for the study by administering the Student Screening Form from the Chinese Education Center to determine language proficiency. No information was included regarding the field testing or norming procedures which were used to develop this measure.

It seems to be an informal assessment measure which was probably developed at a local center.

Analysis of the instrument which was included in the appendix makes its content validity suspect. The test consists of the following sections: I. oral response and comprehension (7 items); II. vocabulary and structure (5 items); III. response to directions (6 items); IV. identifying actions (4 items); and V. identification of letters of the alphabet (7 items). Not only was the number of items used to determine proficiency in each section restricted, but the responses which were called for generally required the same grammatical structure in the response. The rationale for weighting of items was not clear.

A second part of the initial screening consisted of administering the Hoffman Bilingual Scale which the investigator claims indicates whether an individual is monocultural or bicultural. In actuality, this test is a language usage survey which determines which language a bilingual speaker uses in which domain and with whom. The premise of using a language usage survey is that monocultural individuals who are bilingual will use their native language in the home domain as opposed to bicultural individuals who will use both. It was not clear whether this was an intended use of the measure or whether it was one which was attributed to it for the purpose of this study.

Students who scored in the bottom three groups of the language proficiency testing and who also were determined to be monocultural by the Hoffman Bilingual Scale were selected for the study. While Lum claims that students were randomly assigned to one of the two groups, they were not. Four schools participated in the study and students were assigned to either bilingual or ESL classes according to their area of residence. One of the two schools had two classes which were both bilingual. Students who resided in that area therefore did not have an equal chance of being assigned to the treatment or the control group.

Duration: The study was conducted over a nine-month period.

Ages: The children were in first grade. No information was given about their age in years.

Type of Program: Neither program was well explained. The bilingual program consisted of the following elements: one-half hour daily of Chinese language, one-half hour daily of ESL instruction, and one and one-half hours daily of English language activities where English and Chinese were both used. There were no guidelines regarding the amount of time teachers were to spend using English and Chinese in the English language activities. Lum determined through observation and teacher estimate that Chinese was used an average of 50% of the time.

No additional information was provided regarding how the bilingual program of instruction was structured or taught, other than that it was team taught. From the available information, it appears that it was a program of dual language instruction rather than a bilingual program, as the students did not receive content area instruction in their native language. The two bilingual classes were also different in that one had English speaking and non-English speaking students.

The only information given about the ESL program was that it was taught by a single teacher and that it spent one-fourth hour more daily in English language activities than the bilingual method. No information was given about whether the program was a self-contained or a pull-out program. In addition, no information was provided about which methodology was used by the teachers to teach ESL. One may assume that English was taught by the audio-lingual method from Lum's comments about how second languages are acquired in the statement of his hypotheses for the study.

Description: It was hypothesized that non-English speaking Chinese children who are taught by the ESL method will acquire more mature oral English production skills than those who receive bilingual instruction In addition, they will make fewer deviations from standard grammar. The reasoning behind these hypotheses is that since the bilingual method uses a segment of its instructional time for the study of Chinese, the ESL group would perform better due to increased exposure to English. Furthermore, due to the fact that English and

Chinese were both used as media of instruction for the English language activities, the bilingual students would be unable to keep the two languages separate and would speak less grammatically correct.

Design: The independent variables in the study were the two instructional techniques--ESL and bilingual instruction. Lum stated that the teachers were matched, but didn't explain how. No information was given regarding the language proficiency levels in English, their years of experience in teaching, or which methods they used to teach ESL.

After the intervention, which lasted nine months, students were given an interview which was taped and analyzed along the following variables: MLR-mean length of response; M5R-mean length of the five longest responses; SCS-structural complexity score; GrF-measure of grammatical correctness; and NDW-number of different words in the sample. The interview conversation centered around questions about pictures from a commercially prepared language arts series. One teacher used a different set of pictures because she felt that the pictures which the investigator had chosen for the study were too hard for her students. This, plus the fact that the interviews were scored by only one rater, brings the validity and reliability of the scores into question.

Analysis A one-way analysis of variance was computed on the scores of all of the variables both between groups and among classes in each group except GrF. This variable, grammatical difficulty, was analyzed by computing mean scores of its four subtests. A t-test was done to determine differences in female versus male students. Scores were also converted to equivalent age norms for English speaking peers Finally, students were asked to report on the amount of English they used outside of school.

Results: There were no differences between the groups on structural complexity and grammar. On the mean length of response, the mean length of the five longest utterances and the number of different words used, the ESL group had significantly higher scores. Conversion of scores to equivalent age norms showed that the bilingual group performed at the level of three and four year-old English speaking children, while the ESL group performed at a level equivalent to English-speaking children of 3.6-4.6 years of age. There were no differences in the scores of male and female students. There was no significant difference in the scores of classes within the same group. Student reports showed that the ESL group spoke and was spoken to more often in English.

Strengths: The study was conducted with Chinese monolingual students only.

Discussion: One should not conclude from the results of this study that ESL is superior to traditional bilingual education in teaching English to Chinese students, for it has serious flaws in its theoretical base, design, and methodology. Lum never explained what language maturity is or how it is related to or differs from language proficiency or communicative ability. While language maturity may have some relevance to language development of a child's first language, one must question whether it is a realistic construct to measure in relation to second language learners. Likewise, many of the measures which were used to determine language maturity are not related to a speaker's ability to communicate. A second language learner could conceivably have a small vocabulary and produce short sentences which are not grammatically complex and yet be able to communicate well in initial stages of second language acquisition.

Most of the faults in the study are in the areas of design and methodology. With regard to the former, the study was done with a small sample of only fifty-five students. As was previously discussed, group assignment was not random, nor were classes within groups comprised of the same type of students. Criticism may be leveled against the study on methodological grounds in that assessment procedures were not standard for all students in the study as one class was shown a different set of stimulus cards during the oral interview. In addition, the oral interview was administered to each class by its own teacher and was scored by a single rater. Ideally, the same person should have conducted the interview and it should have been scored by at least two raters so that the inter-rater reliability could have been determined. Lum's rationale for using age equivalent scores for English speakers to show differences in the two instructional methods is also questionable. Bilingual children would be expected to score well below their English speaking peers due to the fact that they have not had six years prior to entrance in school to develop English skills. It should be emphasized that while the ESL group scored higher than the bilingual group, it was an average of only 6/10 of one year. This type of result is misleading and is often misunderstood by the general public.

A final methodological problem is that Lum used a language usage questionnaire to determine which group used the most English outside of school at the end of the year. The fact that the ESL group used English more outside of school than the bilingual group is not solely dependent upon the type of instruction they received in school. It is mitigated by the language proficiency levels of those in their home environment as well as the established norms for language use within the community where they reside.

In conclusion, it is hard to explain how this study was included in the Baker/DeKanter report as one of the twenty-eight studies which met their criteria as an acceptable study for inclusion in their report on the effectiveness of traditional bilingual education as an instructional method for limited English-speaking children. Close examination of the study shows that not only must the results of the study itself be questioned, but that its results should not be generalized and used as an indictment against traditional bilingual education.

Reference: Mathews, Tom. An investigation of the effects of background characteristics and special language services on the reading achievement and English fluency of bilingual students. Seatle, Washington: Seatle Public Schools, November, 1979.

Baker de Kanter Review - p. 29-33, Chapter 2.

Available From: Seatle Public Schools, Administrative Services Center, 815 Fourth Avenue North, Seatle, Washington 98109

Name of Study: An Investigation of the Effects of Background Characteristics and Special Language Services on the Reading Achievement and English Fluency of Bilingual Students.

Author and Date: Mathews, Tom; November, 1979.

Location: Seatle, Washington.

Treatment Group: 1747 students identified as "bilingual" (primary or home language other than English), who received special language services (ESL or Bilingual Education).

Comparison Group: 3515 "bilingual" students not receiving special language services.

Duration: One academic year

Ages: All "bilingual" students in Grades K-12. Achievement was measured at Grades 2, 4, 6, and 8.

Type of Program: An investigation of the effects of special language programs (ESL and Bilingual Education) on English reading achievement and fluency of "bilingual" students (students with a primary or home language other than English).

The data base consisted of: achievement test results; amount and type of instructional services; relative English fluency assessments, language background, grade level, and family income. The study attempted to examine the complex relationships among student outcomes, student characteristics, and educational input.

The approach used was the construction of a conceptual model which attempted to analyze the relationship of the following variables:

A. Achievement - Total reading percentile scores on the Metropolitan Achievement Test (MAT) for Grades 2, 4, 6, and 8 were used to measure reading and math achievement in the Spring of 1979. The assumption was made that although special language programs were not designed to teach English reading skills, an academically successful bilingual student would be expected to acquire these skills.

B. Previous Achievement - Total reading percentile scores from the MAT from Fall, 1978.

C. Special Language Services - Subjects of the study were the served and not-served "bilingual" students. Special language service was defined as ESL instruction and/or bilingual instruction.

D. Relative English Fluency - Students were grouped into two categories: low and high fluency.

E. Language Background: All language groups were divided into eight categories, i.e., Chinese, Philippine, Korean, Spanish, Japanese, Somoan, Vietnamese, and Other.

F. Family Income - Students were grouped into two categories as to whether or not they were eligible for free/reduced price lunch.

G. Grade Level - Student grade levels were grouped into two categories: grades 2 and 4 were "lower grades" and grades 6 and 8 were "higher grades."

The author attempted to study the relationship of achievement to the other study variables based on the assumption that all such variables are related to achievement. The results were consistent with the assumption of a direct effect of grade level, income, and language background on the reading achievement of bilingual students. Higher achieving students tended to be from specific language groups, higher income families, and lower grade levels. Consistencies were also found in that higher fluency students tended to come from higher income families and specific language groups and that higher fluency appears to lead to higher achievement. The pattern was reflected by a strong positive relationship between relative English fluency and reading achievement. Interestingly, "bilingual" students' math achievement was at a higher level than other students in the district.

Bilingual students who received special language services tended to achieve at lower levels than those who were not served. The author concludes that this may reflect that lower achieving students were served until their achievement increased, i.e., language services were withheld as achievement scores increased. However, the author also states that these results may indicate the existence of a direct negative effect of special language service on achievement. "Given the limitations of the current data base, it is impossible to determine whether the results are a reflection of the effects of service or whether they merely reflect current program implementation practices." (p. 23).

Strengths: The author is careful to indicate possible design and result problems and/or concerns of the study.

Discussion: Over 60% of the low fluency students scored in the lower three stanines in reading. The author also states that a perfect relationship between fluency and achievement was not obtained, and as a result, some high achieving, low fluency students were identified.

A larger portion of the students from low income families than from high income families scored in the lower three stanines.

Some language groups had higher proportions of low achieving students than other groups. Somoan and Vietnamese contained the highest proportions of low achievers.

Students in Grades 6 and 8 scored lower in reading than students in other grades.

"Bilingual" students who did not receive special language services tended to score higher in reading than students who received such services.

The study attempted to determine the effects of special language programs on reading achievement and English fluency of bilingual students. It came as the result of increased spending in these program areas and attempted to justify the spending with supporting data. However, the author states that the report was unable to reach a conclusion to the question because:

A. The served and not-served groups were not comparable;

B. There was a low sample size in some groups;

C. There was a lack of specific program definitions and objectives;

D. Analyses of achievement levels were limited to "bilingual" students in Grades 2, 4, 6, and 8;

E. When the data used in this analysis were generated, no attempt was made to form comparable served and not-served groups;

F. Much of the study was devoted to an analysis of reading achievement in English, and this outcome is not directly pertinent to objectives of current language programs. Standardized tests utilizing the native language were not available, nor were measures available related to course objectives;

G. Service was delivered to students of perceived greatest need and no service to students not judged in need, therefore service comparisons were unequal;

H. Comparisons of fluency across several groups should be cautious as service was administered mostly to those students who were in low fluent levels, but at times, some students were not served at the parents' request, i.e., service was refused.

In addition to the author's notations of the limitations of the study, it should be pointed out that:

1. Methods of determining fluency were not provided;
2. Descriptions of instructional procedures (ESL and Bilingual Education) were not given;
3. Amount of time provided in ESL instruction and Bilingual Education was not specified;
4. Student achievement related variables, e.g. IQ, age, etc. were not included;
5. Length of time in special language programs was not noted, e.g., attrition, etc.
6. All criteria for placing students in special language programs were not given.

The author should be commended in that, while his study did not answer the question of efficiency of special language programs, limitations and unusual findings were reported. The author concludes that the results indicate one of two things:

1. Students who receive special language services tend to be those with the greatest need for the services, or
2. Special language services inhibit reading achievement.

The study recommends another well developed study which maintains control over the assignment of students to groups to determine which of the two explanations is correct. The author states that none of the conclusions can be defended strongly and recommends another study which maintains control over outcome criteria, program definitions and objectives needed to address questions of service effectiveness.

Reference: McConnell, B.B. Effectiveness of individualized bilingual instruction for migrant students. Unpublished doctoral dissertation, Washington State University, 1980.

Baker De Kanter Review p. 6-10, Chapter 2.

Available From: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, Michigan 48106.

Name of Study: Effectiveness of individualized bilingual instruction for migrant students.

Author and Date: McConnell, B.B.; 1980.

Location: La Grulla, Texas and Pasco, Washington.

Treatment Group: 630 students between the ages of five and nine years of age, classified as having Spanish as their primary language, enrolled in the State of Washington and in Texas. Parents were migrant workers, that is to say, having at least two employers during the year at different locations.

Control/ Comparison Group: The 390 "baseline" comparison students consisted of pre-test scores of students in the bilingual program plus scores of students from surrounding districts. Both treatment and control groups included only students tested between 1974 and 1979, as this was judged to be a period of program stability.

In order to have sufficient numbers in the "baseline" group, students from the neighboring school district were added in with the "baseline" group. Of the 24 possible "baseline" comparisons between the treatment group and students added in from the neighboring school district, five comparisons proved significantly different in favor of greater proficiency of the students from the neighboring district. That is to say, 21% of the comparisons were significantly different; yet, the researcher states "Most ages on all tests, the mean scores from the IBI pre-test and from the comparison school are not significantly different, justifying the combining of the two groups."

Duration: Six years

Ages: 5-9 years of age

Type of program: Bilingual education tutoring program, Spanish language supplemental to the regular education program.

Description: The project consisted of migrant children served through a program of supplemental bilingual education tutoring, operated on an interstate basis between La Grulla, Texas and Pasco, Washington. It had been designed as a supplemental program to whatever schooling was provided by the local school district in each area. For pre-school children at each site, the program offered a day-care program with bilingual curriculum as its educational component. For school-age children, a cooperative arrangement was worked out with the local schools. For example, if kindergarten were offered on a half-day basis by the schools, the children attended the pre-school special project program the other half-day. For children up through the third grade who were in school on a full-day basis, the bilingual program was provided on a release time basis, or after school for an hour to an hour and one-half each day. The Sunnyside School District in the State of Washington served as the fiscal agent for the project. One intent of the program was to provide continuity of teaching staff; therefore, the "teachers" in the program were migrant. (More than half were parents of the children served). The teaching staff consisted of previously untrained and inexperienced bilingual adults. Certified staff used in the project served as supervisors and trainers. (There was one trainer for each seven or eight of the paraprofessional "teachers."

Because release time had to be worked out with the school, the special program was provided mostly in the afternoon, usually with two shifts of children being served in groups of six to eight at a time. On occasion, the project teacher was itinerant, moving from one school to another if migrant children were in different schools. Contrasts in approach between the State of Washington and the Texas programs were many. While release time was provided in the Texas schools; for the most part, in Washington, it was the exception. Therefore, instruction time was arranged with, at most, a half-hour early release for children at the end of the afternoon, and continued for an additional hour after school. Classification of students was based upon their scores on the

Peabody Picture Vocabulary Test. Form A was in English, Form B was a project-developed translation in Spanish. If scores on the Spanish version were higher than scores on the English version, the child was classified as coming from a home in which Spanish was the primary language. This classification was altered on occasion if the teacher were to decide, upon visiting the home, or in further conversations with the child, that there should be a different classification. 75% of all students were classified as having a primary language of Spanish, with 47% of this 75% being Spanish monolingual. The 25% classified as primary language English were nearly all enrolled in the State of Washington program. 77% of these students had English scores below the first percentile for their age level when compared to monolingual English-speaking students.

Teacher training consisted of a brief orientation, followed by students being assigned to them with "intensive training" contiguous to the teacher's work with students over a three-month period. In-service training continued throughout the teacher's employment.

Teachers had greatly differing amounts of training and experience. The initial training methods and curriculum utilized during the first years of the program were "gradually rejected through program experience."

At any given time, the children enrolled represented vastly different prior attendance, as well as language capacity.

Initially in the program, children were taught reading only in English. During the course of the program, specifically in the last years of the program, training staff switched to having children learn to read first in Spanish. This change met with resistance from school personnel and parents, and the project reverted to teaching reading first in English.

Variances in scheduling have been mentioned, but the program in Texas usually lasted one hour for five days a week. In the State of Washington, the program usually lasted ninety minutes for four days a week. The curriculum subject areas of English, math, and reading were covered at least twice each week.

Testing of children occurred during the first thirty days of attendance in the program. These test data served as baseline for comparison to other children who, at the same age, had completed varying periods of program attendance.

Post-testing occurred after 100-day intervals of school attendance. Each month, the "tester" at each site was notified of children who had passed a testing interval in their attendance; thereby triggering the appropriate series of post-tests. Given the nature of the population, i.e., migrant, wide variations in actual length of time between pre- and post-testing existed. "Testers" were "usually part-time personnel responsible for duties other than instruction, such as clerical duties." While results were reported in years of attendance in the program, in reality, years were calculated on the basis of actual days of attendance. For example, one-half year equaled 100 days of attendance, one year equaled 200 days of attendance. Two years were reported as 300-400 days of attendance. Attendance of over 500 days was reported as 3 years.

Strengths: The project was longitudinal in nature, and apparently maintained extensive student records.

Discussion: Results obtained through the application of this study's research design indicate there were significant differences favoring the treatment bilingual program in English vocabulary, Spanish vocabulary, reading levels, and math levels. These differences hold for a variety of analyses and regrouping of data, such as analysis by age, sex, attendance, and so forth.

In the study, a child was considered "balanced bilingual if the score in English vocabulary, upon entry to the program, was within 50% of being as high as the score achieved on Spanish vocabulary." Monolingual was defined as an English auditory vocabulary score so low it could be achieved simply by guessing. Limited bilingual was defined as a child who scored higher than the guessing level on the English vocabulary test, but less than 50% as high as their Spanish vocabulary score.

There were significant differences between the State of Washington subjects and the Texas subjects, with English vocabulary and English reading favoring the State of Washington subjects, and Spanish vocabulary and math favoring the Texas site. Teachers in the State of Washington indicated more use of English, and felt they spoke English as well or better than Spanish, which was the opposite case for teachers in Texas.

The study utilized a time series quasi experimental design model. Given the complexity and array of variables potentially operating within this study, the design suffers greatly for lack of a comparison or control group. The following is an incomplete listing of variables associated with the study that have the potential to influence study results, but were uncontrolled within the research design:

Reference: McSpadden, J. R. Acadiana bilingual-bicultural educational program: final evaluation report, 1980-81. Lafayette, Louisiana: Lafayette Parish School Board, 1981.

Baker de Kanter Review - p. 39-40, Chapter 2.

Available From: Bilingual Resource Center, 7703 North Lamar, Austin, TX 78752

Name of Study: Acadiana Bilingual-Bicultural Educational Program

Author and Date: McSpadden, J. R.; 1981.

Location: Acadiana Perish, Lafayette, Louisiana

	Grade	Number of Students
Treatment Group:	Kindergarten	21
	Grade 1	26
	Grade 2	25
	Grade 3	51
	Total	123
Comparison Group:	Kindergarten	19
	Grade 1	26
	Grade 2	23
	Grade 3	25
	Total	93

Duration: One Year (1980-81)

Ages: Kindergarten through Grade 3

Type of Program: French as a Second Language and transitional bilingual education

Description: The program, encompassing kindergarten through third grade, was implemented in two elementary schools; one public and one parochial. Associate teachers and aides provided all instruction in French in the bilingual classrooms, with the regular teacher teaching only in English in those classrooms. In kindergarten, 1,000 minutes of instruction were provided in French; 1,500 minutes of instruction in English. In Grades 1, 2, and 3, 1,160 minutes of French instruction and 2,700 minutes of English instruction were provided. Further details concerning the instructional methods utilized, classroom configurations, student proficiency levels, etc. are not provided in the document.

Evaluation of the program was accomplished by contrasting the performances between the bilingual instructional groups and comparison groups on a criterion-referenced, program-developed French test, and a standardized English achievement test. The following instrumentation was utilized:

1. French Language Tests of Basic Concepts (a staff-developed, criterion-referenced instrument)
2. Metropolitan Achievement Test (revised)
3. Home Language Survey (revised, project-developed)
4. Community Survey (project-developed)
5. Staff Survey (project-developed)

6. Documentation Forms (various forms utilized for collecting information from the projects)

Comparisons presented in the document included analysis of variance, pre-/post test results of the experimental group in the areas of reading readiness, linguistic structures, mathematics, social living, and social science.

Additional pre-/post-test comparison by analysis of variance and covariance was accomplished between experimental and comparison groups in language arts and mathematics.

No significant differences were noted between pre- and post-test results for the experimental group.

It should be noted in the comparisons of experimental and control groups, it is reported that the control group scored significantly higher on pre-testing; therefore, post-test comparisons of experimental and control groups must be questioned. It is reported, however, that in the area of language arts, post-test comparisons between experimental and control groups indicate that at the kindergarten level, there was a significant difference in favor of the experimental group in the area of word sounds. In the area of mathematics, the only significant difference reported was at Grade 3 where control group students performed significantly better than experimental students. A summary chart presented in the report of the differences by subject and by grade between experimental and control groups reflects either typographical or editorial errors, as the discussion does not coincide with the chart.

Strengths: Apparently, an independent external evaluator was employed to accomplish the evaluation.

Discussion: The study as reported displays the following types of problems based upon either missing data and/or design analysis problems:

1. There is no data concerning sample selection; therefore, it is unknown whether experimental and control groups were randomly selected or selected on other criteria.
2. No information is provided concerning the specific types of instruction, other than it is known that associate teachers, rather than regular teachers, provided all instruction in French, raising questions of their qualification and quality of instruction provided in the second language.
3. It is indicated that experimental and comparison groups were significantly different on pre-testing; therefore, results reported for post-testing must be questioned.
4. It is not known whether there were any efforts to control for known variables associated with achievement, that is to say, I.Q., time on task, handicapping conditions, etc.
5. There is no information provided relative to the language proficiency of students in either experimental or control groups.

Reference: Moore, F.B. & Parr, G.D. Models of bilingual education: comparisons of effectiveness. Elementary School Journal, 1978, 1979(2), pp. 93-97.

Baker de Kanter Review - p. 37-38, Chapter 2.

Available From: Perry-Castaneda Library, University of Texas at Austin

Name of Study: Models of Bilingual Education: Comparisons of Effectiveness.

Author and Date: Moore, F.B. & Parr, G.D.; 1978.

Location: Four elementary (Title VII) schools in West Texas

Treatment Group: 130 Spanish-dominant students

Control Group: 77 English-dominant students

Duration: School Year (6-8 months)

Ages: Grades K-2

Type of Program: Non-random assignment of students to one of four different programs:

A. Maintenance - at least 50% Spanish spoken

B. Transitional - Spanish used as needed (starting with about 50% and decreasing)

C Minimal - no more than 20 minutes of Spanish spoken a day.

D. Monolingual - no Spanish spoken in the classroom.

The number of students assigned to each group was not given.

Description: The goal of the study was to measure and compare the effectiveness of the four programs used in four elementary schools.

The teachers administered pre-tests in the Fall and post-tests in the Spring (6-8 months apart).

Assessment Instruments:

K. Tests of Basic Experience

1. Comprehensive Test of Basic Skills
   Comprension del Lenguaje
   Primary Self-Concept Inventory

2. Comprehensive Test of Basic Skills
   Pruebas de Lectura
   Cultural Attitude Scale

Strength: "Analysis of covariance was used for all measures to correct statistically for the non-random assignment of subjects." Examples of variables covaried are: sex, school, teacher, teacher competency, and aide competency.

Discussion: Mobility of the population due to seasonal work was discussed, but attrition was not dealt with in the paper. Other student achievement related variables such as IQ, handicapping conditions, etc. were neither controlled nor addressed in this study.

The kindergarten children in monolingual, minimal, and maintenance scored significantly higher on Tests of Basic Experience than did those in the transitional class. Maintenance and transitional were very similar, and therefore program differences could not be claimed. Furthermore, there were no significant differences reported in the two extreme programs of monolingual and maintenance.

The monolingual scored significantly higher on the reading scale and the language scale of the *Comprehensive Test of Basic Skills*.

Girls scored significantly higher than boys on reading, language, and self-concept. Teacher and aide competency, as rated by the research director, showed an inverse significant relationship to reading and language tests on the *Comprehensive Tests of Basic Skills and Cultural Knowledge*. English-dominant pupils scored significantly lower than Spanish-dominant or bilingual pupils on *Comprension del Lenguaje*.

Conclusions were: "There was no compelling evidence that (maintenance and transitional) these two approaches differed." Measures of Spanish language or reading did not show one program better than another. Only in reading and language achievement in English did one program show a significant difference and that was the monolingual, scoring higher than any of the other three bilingual programs. A possible explanation for no difference is the extremely short (6-8 month) time span pre-/post-test. It would be unusual for significant differences to emerge on the tests selected within this time span of testing. The observation of an inverse relation of teacher competence to achievement casts doubts upon many aspects of the experiment and further clouds conclusions relative to the instructional models investigated.

Reference: Olesini, Jose. The effect of bilingual instruction on the achievement of elementary pupils. Unpublished doctoral dissertation, East Texas State University, 1971.

Baker de Kanter Review - p. ???

Available From: Latin American Collection, University of Texas at Austin (microfilim).

Name of Study: The Effect of Bilingual Instruction on the Achievement of Elementary Pupils.

Author and Date: Olesini, Jose; 1971.

Location: Harlandale Independent School District, San Antonio, Texas

Treatment Group: Thirty third-grade students randomly chosen from Harlandale ISD, a district with a broad student cross-section, urban, rural, migrant. All thirty had had two years of bilingual instruction.

Control Group: Thirty randomly chosen third-grade students from Harlandale ISD with no previous bilingual instruction.

Duration: Fall to Spring in one school year (7 months)

Ages: Third grade

Type of Program: Transitional bilingual education

Description: The goal of the study was to evaluate the achievement of two composite groups of bilingual students in: vocabulary, reading, spelling, language, arithmetic computation, and concepts.

The Otis Quick Scoring Mental Ability Test was given to both groups as a control for IQ, and there was no significant difference. In

October, the Metropolitan Achievement Test, Battery A was given, and Battery B was given in April.

There was a significant difference in vocabulary, reading, language, and arithmetic concepts; the bilingually educated group faring better. Olesini concluded that there was no significant difference between the two groups in spelling and arithmetic computation.

Strengths: A strength of the study was the control for mental ability. Also an effort (probably successful) was made to locate an adequate cross-section of different socio-economic levels that also adequately represented bilingual populations and populations were chosen randomly.

> "Sixty third-grade children were chosen at random. Thirty of these were chosen from three sections that were instructed bilingually for at least two years... The children in these classes totaled seventy-two. They were assigned numbers and the odd-numbered students were selected until the desired numbered group of thirty was obtained."

Likewise, three sections of no previous bilingual instruction students were assigned numbers and the even-numbered students were selected making thirty.

The chronological ages of experimental boys was 109 months, and the control boys, 108 months. The experimental girls were 104 months, and the control girls were 108 months.

Discussion: The chief limitation of this study was the short time frame - seven months from pre-test to post-test. Additionally, an assumption was made that Harlandale ISD was representative of all bilingual school districts.

The results of Olesini's study were: that bilingual students, both sexes, achieved greater gains in academic curricula when instructed with bilingual methods. He determined (p. 51) that

> It can be concluded, either from the average grade placement achievement for the experimental groups, or the individual achievement in the areas of curricula that bilingually instructed bilingual children make a significantly greater academic achievement than a similar group of children instructed by the traditional method."

Reference: Pena-Hughes, Eva & Solis, Juan. abcs. McAllen, Texas: McAllen Independent School District, Mimeographed, 1980.

Baker de Kanter Review - p. 71-72, Chapter 2.

Available From: Authors of paper, McAllen Independent School District, McAllen, Texas

Name of Study: abcs

Author and Date: Pena-Hughes, Eva & Solis, Juan; 1980.

Location: McAllen, Texas

Treatment Group: 78 students (Grades K-1)

Comparison Group: 78 students (Grades K-1)

Duration: Not reported, but possibly 1 or 2 years.

Ages: Grades K-1 - results for those two grades were reported.

Type of Program: Labeled an immersion system, but incorporated elements of the following three types:

1. Total - target language (English) used exclusively (except for 50 minutes per day).
2. Partial - half day each language, never mixed.
3. Mixed - Language mixed as necessary, i.e., answer in language of students' choice.

Description: Elements of all three approaches were used.

"The goal of abcs is to have as many native Spanish-speaking students reading at grade level in English by the end of Grade 1." Beyond the basic curriculum was a time slot for teaching Spanish through a basic reading process.

This program "is characterized by the RAPID, NONTRAUMATIC ACQUISITION OF ENGLISH and SOPHISTICATION OF THE NATIVE LANGUAGE."

The focus was on "acquisition" being naturalistic and sophistication," i.e., not losing the native language while acquiring another language.

"abcs" assumes: 1) quick, easy acquisition of language by children; 2) learning a second language is not qualitatively different from learning a first; 3) interference between the two languages is "the least influencing factor in acquiring the second language."

Strength: The strength of the program was that teachers were described as maintaining a positive attitude that the students could acquire the target language.

Discussion: There is no discussion in the paper about the criteria for assigning the students to the control and experimental groups nor the duration of the study. Time interval for pre-/post-testing was not provided. There was no explanatory narrative on the results.. The results of "abcs" are listed on twelve one-way analyses of variance source tables.

The English pre-test scores on the Language Assessment Scales were utilized for pre- and post-testing; English and Spanish LAS were administered. Significant differences between experimental and control groups existed on the pre-test Spanish LAS measure. Significant differences between experimental and control groups were obtained both on the English and Spanish post-test LAS. Experimental groups had higher means on both the pre- and post-tests, there were no statistical or other controls reported for the fact of non-equivalent groups.

Reference: Plante, A.J. A study of the effectiveness of the Connecticut 'pairing' model of bilingual-bicultural education. Hamden, Connecticut: Connecticut Staff Development Cooperative, 1976.

Baker de Kanter Review - p. 11-12, Chapter 2.

Available From: ERIC Document Reproduction Service, ED 125 260.

Name of Study: A Study of the Effectiveness of the Connecticut "Pairing" Model of Bilingual-Bicultural Education.

Author and Date: Plante, A.J.; 1976.

Location: Cheever and Columbus Schools, City of New Haven School System, New Haven, Connecticut.

Treatment Group: Forty-five students were randomly selected from a group of seventy-six Spanish-dominant students living within the attendance areas of the Cheever or Columbus Schools. At the end of the two-year study period, the treatment group was reduced to fourteen boys and seventeen girls. This reduction was caused by study children moving out of the area.

Control Group: Twenty-seven students were randomly selected from a group of seventy-six Spanish-dominant students living within the attendance areas of the Cheever or Columbus Schools. At the end of the two-year study period, the control group was reduced to ten boys and twelve girls. This reduction was caused by study children moving out of the area.

Duration: Two years (1973-1975)

Ages:	Grades 1 and 2.
Type of Program:	Pairing model of bilingual-bicultural education.
Description:	The pairing model consists of one native Spanish-speaking teacher who teaches basic skills in Spanish and an early childhood trained English-speaking Anglo teacher who teaches English at the same time, beginning with an aural-verbal approach. When an English oral vocabulary is sufficiently developed in individual children, instruction in the reading and writing of English is then initiated. The instructional organization of the pairing model is a diagnostic-prescriptive approach with both Spanish and English resources being available.

Reference Goals and Questions

To determine the effectiveness of the "pairing" model in improving the actual success of Spanish-dominant children who are typically classified as low achievers.

1. Can a carefully designed model of bilingual-bicultural education improve the reading skills achievement of Spanish-dominant elementary school children?

2. Can a carefully designed model of bilingual-bicultural education improve the basic skills (arithmetic and language arts) achievement of Spanish-dominant elementary school children?

3. Can a carefully designed model of bilingual-bicultural education maintain or improve the self-concept of Spanish-dominant elementary school children?

Strengths: Care was taken that experimental and control groups were similar. Statistical methods and findings were reported in detail.

Discussion: Spanish dominance was determined by students' performance on the Oral Vocabulary subtests of the Inter-American Test of General Ability, which was administered to all Spanish surnamed children successfully completing the kindergarten and first grade in the spring of 1973. Students were generally described as Puerto Rican immigrants from low income families. At the beginning of the report, the Cheever and Columbus Schools were described as inner city schools that serve a "high concentration of welfare families," having a majority language and culture different from the minority group of Spanish-speaking students. Later the report described the school environment as having "a Puerto Rican pupil population of approximately fifty percent, growing larger," and being located "in one of the poorest sections of the city of New Haven." Although students from both the Cheever and Columbus school attendance areas were in the experimental and control groups, the experimental classroom was located only on the Columbus campus. Students who were in the Cheever school attendance area were bussed to Columbus if they were chosen for the experimental group, but remained at Cheever if they were selected for the control group. It was not noted in the report what percentage of the control group received instruction at Cheever school and what percentage received instruction at Columbus school, nor were any differences noted between the two schools. An analysis of pre-study and post-study characteristics revealed only minimal changes. Children compared at the end of the study had the same quality of similarity which existed at the beginning of the study. If a chance advantage did exist as a result of the loss of students over the two-year period, it was felt that it favored the control group students. The Metropolitan Achievement Tests were administered to assess basic skills achievement. It was concluded that the pairing model did increase the reading achievement of Spanish-dominant elementary school children at a statistically significant level (.10 to .005). The model increased English reading achievement at all grades; the increase was statistically significant at the second-grade level. Arithmetic and language arts skills were also improved in comparison with those of children in typical classrooms. The report also indicated that the pairing model did enhance the development of a positive self-concept in Spanish-dominant children, who exhibited less negative behavior than their control group counterparts.

Throughout the report, it was clear that many assumptions were made about the Spanish surnamed students in general. Not only were they described as immigrants and economically, culturally, and linguistically deprived, but were also described as low achievers and having poor self-images. They expected little of these students other than an early withdrawl from the educational process. If the author's descriptions reflected a general attitude of the educational community, there is the question of what effect this had upon those students. No comparison was made between the experimental group and total school norms or national norms regarding achievement.

Instructional procedures are not adequately defined to allow conclusions relative to effects of or amount of time devoted to instruction by the bilingual teacher and the Anglo English-speaking teacher. Determination of oral vocabulary development is not specified, but was critical according to design, for instruction was alternated to the second language at some unspecified point of English language development.

Reference: Ramos, M., Aguilar, J. & Sibayan, B. *The determination and implementation of language policy*. Quezon City, Republic of the Philippines: Phoenix Press, 1967.

Baker de Kanter Review - p. 24-28, Chapter 2.

Available From: Oceana Publications, Inc., Dobbs Ferry, New York, 10522.

Name of Study: The Determination and Implementation of Language Policy.

Author and Date: Ramos, M., Aguilar, J. & Sibayan, B.; 1967.

Location: Republic of the Philippines

Iloilo Experiment I (1948-1954) and Iloilo Experiment II (1961-1964) - Iloilo Province

Rizal Experiment (1960-1966) - Rizal Province

Treatment Groups: Iloilo Experiment I - 82 Hiligaynon-speaking pupils
Rizal Experiment - approximately 600 Tagalog-speaking pupils
Iloilo Experiment II - approximately 300 Hiligaynon-speaking pupils

Control Groups: Iloilo Experiment I - 82 Hiligaynon-speaking pupils
Rizal Experiment - approximately 600 Tagalog-speaking pupils
Iloilo Experiment II - approximately 600 Hiligaynon-speaking pupils.

Duration: Nine years total
Iloilo Experiment I - six years
Rizal Experiment - six years
Iloilo Experiment II - three years

Ages: Grades 1-6

Iloilo Experiment I - Grades 1-6
Rizal Experiment - Grades 1-6
Iloilo Experiment II - Grades 1-3

Type of Program: Iloilo Experiment I - bilingual education/immersion
Rizal Experiment - bilingual education/English as a Second Language
Iloilo Experiment II - trilingual education/immersion

Description: This book is divided into three parts leading up to a state education plan of language instruction. In the first part, the author, Maximo Ramos, examines the language policies of the Dutch, British and French colony states in South and Southwest Asia, i.e., Burma, Ceylon, Cambodia, Laos, Vietnam, India, Indonesia, Malaysia, and Pakistan. The study focused on these Asian states because, like the Philippines, they became independent after World War II and were faced with urgent problems arising from a multiplicity of languages within their borders. The study of these various language policies resulted in a statement of implications for the Philippines:

1. Need to put an end to prejudice on the part of Filipinos against their own native languages.

2. Need for a clearer understanding by more people of the nature of language and the psychology of language learning.

3. Need for making haste in re-adjusting a language situation.

4. Need to keep English.

5. Need to keep in mind the value of language as a unifying force.

6. Needless waste in having to teach Spanish in the schools.

7. Soundness of recent changes in the media of instruction used in the Philippine primary schools from English to the vernacular.

Jose Aguilar, the author of the second part of the book, reviewed numerous research findings and the role they played in the determination of language policy. Of the numerous studies reviewed, three dominated his discussions. These were Iloilo Experiment I, the Rizal Experiment, and Iloilo Experiment II.

The Iloilo Experiment I

This six-year study took place in the Iloilo Province from 1948 to 1954. It was characterized by the preparation of teaching materials and by a motivation to commit ideas to trial. The study showed the possibility of better English learning when based on a well-supported study of Hiligaynon, the vernacular of Iloilo Province.

A massive drop-out rate and a nation-wide emphasis on the "3 Ls" - language, literacy, and living - led to the study of the educational and social value of teaching Hiligaynon, followed by the teaching of English, with particular reference to the lower grades. The study was designed to have children followed from Grades 1 through 6, giving instruction to the "experimental group' in Hiligaynon in Grades 1 and 2, and in English from Grades 3 through 6, and to the "control groups" in English from Grades 1 through 6. A purpose of the design was to find out in Grades 1 and 2 which of the two languages was more effective in teaching children, and also to determine which of the two beginning languages, Hiligaynon or English, was more effective in learning English from Grades 3 to 6.

Originally, there were 758 pupils in the "treatment group" and 1164 in the "control group." Because of the incidence of drop-outs, these numbers decreased from year to year. At the end of the six-year study, 232 remained in the "treatment group" and 301 in the "control group." Of these, only 82 pupils from each group could be matched. Seven elementary schools were selected for the experimental group; one in the city, three in a farming area, and three in a fishing area. In each of the last two areas, three schools classified as low, average, and high in reference to economic conditions were chosen; the school in the city was average. Seven control elementary schools similarly located and similarly classified were selected.

The Superintendent of Schools for Iloilo Province took charge of organization, writing and supplying of teaching materials, and the conduct of the experiment in the province. As an added measure in the control of variable factors, each of three district supervisors involved was assigned a control and an experimental school or class. On the basis of educational qualification, civil service eligibility, experience and efficiency, an attempt was made to equate teachers and principals for the experimental and control groups.

Teachers were trained to teach Hiligaynon and were supplied with supporting materials. The teaching materials for Grades 1 and 2, experimental classes, consisted of Hiligaynon translations from English of the course of study, texts, readers, and various teaching aides. The teaching program, the time allotment of each school subject, the length of the school day, the length of the experimental period, and the length of the school year were the same for both groups. The classes in both groups were approximately the same size, and were operating under the two-single-session plan. The school equipment and facilities were also noted as approximately the same in both groups.

The Division of Measurement and Evaluation of the Manila Office of the Bureau of Public Schools helped plan the experiment and drew up a program of testing for the six-year period, including the construction and administration of tests. The results of the Philippine Mental Ability Test, Forms I, II, and III, chronological age, and school attendance were used as control factors in annually equating children from control groups and experimental groups. The Philippine Achievement Tests were used to assess educational achievement in reading, arithmetic, language, and social studies. At the end of the six-year period, pupils in the experimental group were statistically superior (.05) to the pupils in the control group in the area of social studies. In arithmetic and in reading, the pupils in the experimental group had a significant edge over pupils in the control group. The pupils in the experimental group were better adjusted, personally and socially, than the pupils in the control group. They were more dominant, extroverted, emotionally and socially mature than the control pupils. The difference between the two groups in emotional stability and emotional maturity was significant at the .05 level.

Rizal Experiment

Part I of the experiment was designed to yield information helpful in deciding when reading activities should be first introduced in the teaching of English as a Second Language in Tagalog-speaking areas. Part II was designed to yield information helpful in deciding when English should be introduced as the medium of general classroom instruction.

Part I included about 600 pupils in experimental and control groups. The experimental group began reading activities in Grade 1, while the "comparison group" began reading activities in Grade 2. About half of each group first used English as the medium of instruction, beginning with Grade 3, while the other half of each group first used English as the medium of instruction beginning with Grade 5. Part II of the experiment involved approximately 300 pupils in each group. All three groups had begun reading activities in Grade 1. One experimental group first used English as the medium of instruction in Grade 1, while the other first used English as the medium of instruction in Grade 3. The comparison group didn't use English as the medium of instruction until Grade 5.

Each class was made up of fifty students with only one experimental class in any given school. Thirty communities were selected, six in each of the following categories: urban, semi-urban, farming, fishing and cottage-industry. Schools were controlled for quality of facilities and randomly assigned to experimental treatments. Teachers were controlled for equivalence on the basis of professional competence, experience, and level of education completed.

After the teachers and the first-grade experimental classes had been assigned to each of the groups, a ten-day seminar was conducted for them prior to the opening of school. A similar seminar was conducted for the second-grade teachers before the opening of school in 1961. In these seminars, demonstrations and discussions were provided of the methods to be employed in teaching English as a Second Language. For third- and fourth-grade teachers, a regular eight-week undergraduate course on second language teaching was offered exclusively for them at the Philippine Normal College. Teachers of the fifth-grade classes attended a six-week seminar on the teaching of English as a Second Language, and for the teachers of Grade 6, it was a four-week training course. During the course of each school year, monthly conferences for teachers of the experimental classes were held, at which demonstrations and procedures were given and problems discussed. Supervision of subject matter areas was provided by regular supervisory staff of Rizal.

A language aptitude test was constructed by the Bureau of Public Schools and administered to pupils in June, 1960 to assess proficiency in English and Tagalog. In addition to language proficiency, chronological age, daily school attendance, and socio-economic level were used to assure equivalence between experimental and comparison groups.

The following tests were prepared and administered by the Bureau of Public Schools in April, 1964: English Language Test, English Reading Test, and Tagalog Reading Test. Other tests were written in English and translated into Philipino to produce as nearly as possible an equivalent in that language. Three versions of each test were then printed; an English version, a Philipino version, and a special version that had parallel columns on each page that had the same items in English and Philipino, which was referred to as the bilingual version. These tests were in the areas of social studies, health and science, and arithmetic. Parallel forms of these tests were administered again in April, 1966.

In this experiment, conditions differed from those in the Iloilo Experiment I; teacher training was concentrated in English, neglecting the home language, Tagalog. Tagalog teaching materials were anchored and made equivalent to the English materials. Under these differing conditions, the group with English as the medium of instruction showed, at the end of Grade 6, achievement in English proficiency, and in social studies, health and science, and arithmetic significantly greater (.05 level) than the achievement of the groups that used the Tagalog medium in Grades 1 and 2, or in Grades 1 through 4, regardless of the language used for measuring achievement. At the end of Grade 4, when tested in English, the group instructed in English was highest in achievement in language, reading, social studies, health and science, and arithmetic computation. However, for arithmetic problems, the group instructed in Tagalog attained the highest level of achievement. In the Tagalog version of the tests, the three groups showed about the same proficiency levels in reading, but the group with Tagalog as the medium of instruction obtained the highest achievement levels in social studies, health and science, and arithmetic problems. However, for arithmetic computation, the highest level of achievement was demonstrated by the group with English as the medium of instruction.

For Part I, there was no statistically significant difference between experimental and comparison groups, indicating that proficiency in English is not dependent upon when reading instruction is begun.

In the second part of the experiment, findings were statistically significant, at the .05 level, indicating that proficiency in English is directly related to the number of years that English is the medium of classroom instruction. The data with respect to achievement in Tagalog also indicated that pupils become about equally literate in their vernacular regardless of what language is used as the medium of instruction. Results of the study also indicated that training materials in a vernacular should be developed independently of what is prepared for English instruction.

Iloilo Experiment II

A three-year project was conducted in the province of Iloilo from 1961 to 1964 by the local school system and the Philippine Center for Language Study, with support from the Manila Office of the Bureau of Public Schools. The Iloilo Experiment II, like the Rizal Experiment, was originally designed to find answers to the problem of designing the language curriculum for the elementary grades. As the experiment progressed, it provided an environment for testing the theory of second language teaching in its application to the six-year Philippine elementary school, with the problems of teaching the vernacular, Philipino, and English. The major purpose of this study was to determine whether two second languages should be introduced at the same time, or one after the other.

The following was provided: a full-time coordinator, part-time research consultant, test administrators, training and supervision of teachers of the experimental classes, special curriculum materials, tests and test supplies. Achievement tests for Grade 3 were supplied by the Bureau of Public Schools. Teachers' guides and pupils' books were prepared in the Philippine Center for Language Study by curriculum writers for the Bureau of Public Schools. The Center contributed editorial, technical, and financial assistance.

All classes in the experiment were located in a Hiligaynon-speaking area and used that language as the medium of instruction in Grades 1 and 2. In Grade 3, the medium of instruction was English. The classes were assigned at random to three groups, each of which followed a different language scheme in Grades 1 and 2. The pupils in group 1 studied English for 30 minutes a day and Philipino for 30 minutes a day in each of Grades 1 and 2. The pupils in group 2 studied English for 30 minutes a day in both Grades 1 and 2 and Philipino for 60 minutes a day in Grade 2 only. The pupils in group 3 studied Philipino for 30 minutes a day in Grades 1 and 2 and English for 60 minutes a day in Grade 2 only. All classes studied English for 60 minutes a day and Philipino for 50 minutes a day in Grade 3.

Three sets of seven schools each were formed in such a way as to make the three sets essentially the same with respect to types of communities represented, school facilities, and teacher background, and were randomly assigned to the three schemes of language learning. The 21 experimental classes were each made up of 22 boys and 22 girls randomly selected from all of the 1961 applicants for entrance to Grade 1. Attrition was implied present but was not specified. Schools were selected from communities that were characterized as primarily fishing communities, agricultural communities, semi-urban communities, or a mixture of these types. All experimental classes were in the central schools of the districts selected.

With the form used in the Rizal Experiment for rating school facilities, the quantitative ratings of each of the 21 schools were obtained for the first and third years of the experiment. Teacher background indexes were also obtained by summing the number of years of paid experience, official efficiency ratings, and quantitative representations of professional training. Since the original 21 classes were promoted from grade to grade as intact groups, the pupils in each class had the same three teachers during the three years of the experiment.

Before the experiment began in June, 1961, a ten-day seminar was held for teachers of the experimental classes in Grade 1. The purposes of the experiment, the meaning and principles of second language teaching, and the use of the teaching guides were explained. The content of the curriculum in each of the three experimental groups was outlined. Demonstrations of teaching techniques were presented and discussions of second language teaching were conducted. Teachers who taught the experimental classes in Grades 2 and 3 took an undergraduate course in the Teaching of English given exclusively for them at the Iloilo Normal School during the summer preceding their assignment to experimental classes. Monthly conferences of teachers of the experimental classes were held during the course of each school year, where demonstrations of teaching procedures were provided and discussions of problems were encouraged.

The Philippine Center for Language Study constructed the English Proficiency Test and the Philipino Proficiency Test specifically for the project. Each test had four parallel forms and was comprised of four subtests: listening comprehension, oral expression, reading comprehension, and written expression. In June of 1961, the tests were administered by a selected group of 42 teachers who were instructed in the procedures involved. In 1962, 1963, and 1964, the tests were given by eight specially trained examiners, four of them selected for oral proficiency in English, and four of them selected for oral proficiency in Philipino. To insure uniformity of administration, each examiner gave the same parts of the tests in all 21 classes. The achievement tests in the areas of social studies, health and science, and arithmetic were constructed by the Bureau of Public Schools and were administered in April, 1964 by three supervisors from the Research, Evaluation, and Guidance Division of the Bureau of Public Schools. They were assigned by three Division Supervisors of Iloilo and the eight trained examiners.

Since the immediate purpose of language teaching was to permit pupils to learn subject matter content, the achievement tests were administered to all pupils in groups 1, 2, and 3 at the end of Grade 3. To make sure that the language used in the test would not prevent conclusions being drawn about differences in achievement, each test was prepared and administered in three versions: English, the medium of instruction in group 3; Hiligaynon, the native language used as the medium of instruction in Grades 1 and 2; and bilingual, with English and Hiligaynon in parallel columns on each page. Near the end of Grade 3, the boys and girls in each class were assigned separately and at random to either the English, the Hiligaynon, or the bilingual versions of the tests in social studies, health and science, and arithmetic.

Statistical procedures were not clearly outlined, but differences among the adjusted mean scores of the three groups on all four measures of English proficiency were significant at the .01 level leading to the conclusion that three schemes of language teaching resulted in different levels of achievement. Pupils in group 1 (30 minutes of English and 30 minutes of Philipino instruction per day) obtained the highest level of mean scores in all four phases of English.

The differences among the adjusted mean scores were significant at the .01 level for oral expression and for reading comprehension in Philipino. In both instances, the pupils in group 1 obtained the highest means. The differences among the adjusted means for listening comprehension in Philipino were significant at the .05 level.

It was concluded that when both Philipino and English are to be taught as second languages in Grades 1, 2, and 3 to children whose native language is Hiligaynon, it is better to begin the teaching of both English and Philipino as second languages in Grade 1 than to begin the teaching of one of them in Grade 1 and the other in Grade 2. It was also concluded that it is better to spread the study of English over all three grades than to limit it to only Grades 2 and 3.

The achievement tests given in English, Hiligaynon, and bilingually at the end of Grade 3 found significant differences (at the .01 level) on the arithmetic test given in Hiligaynon. The major conclusion to be drawn from the data was that the achievement of the pupils at the end of Grade 3 in social studies, health and science, and arithmetic was not appreciably affected by the scheme of language study provided in Grades 1 and 2.

In the third and final part of the book, Bonifacio Sibayan outlines the implementation of the various language policies by the Republic of the Philippines from 1899 to 1966. The results of each policy are described along with their implications for succeeding policies with emphasis upon the most recent period of implemented language policy, 1957 to 1966.

Strengths: This book represents a well-documented history of language instruction in the Republic of the Philippines from the beginning of the twentieth century through 1966.

Discussion: Ramos, Aguilar, and Sibayan's book should be considered as no more than a secondary source as the information presented about these three experiments is not complete for in-depth study. In general, these findings suggest that students in Grades One through Six learn better if language instruction is bilingual or trilingual. with the medium of instruction being the vernacular of the area. At the end of the book Sibayan noted:

> after a decade of implementation (1957-1967), not only must the Revised Educational Program be re-examined, but the results must also be evaluated. For example, vernacular education should be carefully evaluated to find out the reasons for the dismal failure of the program as indicated in the literacy surveys conducted by the Bureau of Public Schools from 1960 to 1962. The survey showed that in four years of education, (the first two of which were in the vernacular), the literacy rate measured with an admittedly very easy test was 53.28 per cent in the vernacular, 36.97 per cent in Philipino, and 28.99 per cent in English. It looks like we are not succeeding in any of the three languages. Was not one of the claims (hence, cause for great expectations) of vernacular teaching that of literacy at least in the native language and that two years of instruction in it was sufficient to guarantee this literacy? It is quite understandable that the rate of literacy in a foreign language is only 29 per cent, but in the native language should not literacy be close to 100 per cent?

The Revised Educational Program was based on considerable research findings which suggested that such a plan would be most effective. But these policies have undoubtedly been revised and the authors do not cover subsequent language policies and findings during the fifteen years which have elapsed since the book was written.

Reference: Stebbins, L.B., St. Pierre, R.G., Player, E.C., Anderson, R.B., & Crum, T.R., *Education as experimentation: a planned variation model. volume IV-A and IV-D, evaluation of follow through*. Cambridge, Massachusetts: ABT Associates, 1977.

Baker de Kanter Review - p. 48-50, Chapter 2.

Available From: ABT Associates, Inc., 55 Wheeler Avenue, Cambridge, Massachusetts.

Name of Study: SEDL (Southwest Educational Development Laboratory)-sponsored Follow Through Programs

Author and Date: Stebbins L.B., et al; 1977.

Location: SEDL-sponsored sites:

Philadelphia, Pennsylvania
Los Angeles, California
Tulare, California
St. Martin Perish, Louisiana
San Diego, Texas

Treatment Group: 262 Follow Through Students (Cohort III) (Appendix Tables 14-28)

Control Group: Non-follow through students

Duration School Years 1971 - 1975 (Cohort III)

Ages Kindergarten to third-grade students

Type of Program: Compensatory Education Follow Through Program

Description: Follow through (FT) students registered in all sites routinely provided background information. FT students were evaluated in the broad categories of: (1) Basic Skills (measured by four subtests of the Metropolitan Achievement Test); (2) Cognitive Conceptual Skills (measured by Raven's Progressive Matrices Test); (3) Affective-Cognitive (measured by the Coopersmith Self-Esteem Inventory and Intellectual Achievement Responsibility Scale (IARS).

Non-follow through (NFT) were used for comparative purposes. Their progress/attitudes were measured on the same scales as the FT students. Covariates for the study included: full kindergarten, WRAT (Wide Range Achievement Test), family income, missing income, occupation, missing occupation, entry age, ethnicity (black v. white), ethnicity, (white v. other), sex, first language, Head Start membership, some pre-school membership, WRAT, Head Start, Total Reading, Total Math, spelling, language, Raven's, Coopersmith, IARS, (+), IARS, (-), word knowledge, reading, math computation, math concepts, math problem-solving, Language Part A, Language Part B.

Results of the study show:

1. "At entry to school, only 30% of the predominantly Spanish children scored at or above the NFT median while at the end of the third grade, 50% achieve such scores." The SEDL children made substantive progress. (p. 168 - IV-D).

2. "SEDL sponsor shows rather dramatic gains in mathematics" but the FT median fell at a level below that of their appropriate grade level. (p. 201 - IV-D).

3. "The fact that the SEDL program appears to produce as much progress in reading and in most math areas as is realized in the com-

parison group, is a very favorable finding. (p. 67 - IV-D).

Appendix 14-28 shows the following SEDL FT student characteristics:

1. 63.9% of the total was bilingual;
2. 34% of the group was Black;
3. First language of 40.3% was Spanish;
4. Approximately 50% of data for family size, family income, mother's education, highest occupation in the home, was missing;
5. 5.2% of families had income over $9,999;
6. Mean age of students entering the program was 64.5 months.

Strengths: A mass of longitudinal data appears to be available for study. A cohort analysis research design is employed with results analyzed via ANCOVA. Statistical adjustments were made to scores as needed.

Discussion: Although the analysis of data resulting from seven school years presents the promise of critical differentiations between FT and NFT students being possible, such does not materialize. Several conclusions/facts make this true:

1. Data, statistical information and design particulars are presented in such an unorganized manner that sensible conclusions are nearly impossible. Information is not presented as a body relating to FT cohorts and sponsors.

2. "Approximately 70% of the FT and NFT children who were tested in the kindergarten years of Cohort III-C are not present in the analysis sample." (p.82) Remaining students were felt to be representative of differences in levels of covariates of income and pre-test scores. Other covariates are not analyzed, however, to determine evenness of distribution.

3-a NFT students generally receive assistance from Title I and other programs, so FT has not been compared to the absence of compensatory education.

3-b In some locations there is evidence that some NFT classrooms picked up FT techniques and materials from enthusiastic FT participants, thus modifying the practical distinction between treatment and control. (P. 161 - IV-D).

Reference:	Stern, C. Final report of the Compton unified school district's Title VII bilingual-bicultural project: September, 1969 through June, 1975. Compton, California: Compton Unified School District, September, 1975. Baker de Kanter Review - p. 53-54, Chapter 2.
Available From:	Bilingual Education Department, Compton Unified School District, 604 South Tamarind Avenue, Compton, California 90220.
Name of Study:	Final Report of the Compton Unified School District's Title VII Bilingual-Bicultural Project: September, 1969 through June, 1975.
Author and Date:	Stern, Carolyn; September, 1975.
Location:	Compton, California.
Treatment Group:	A) Year One--37 Mexican-American students in bilingual classrooms (K-1) B) Year Two--27 Mexican-American kindergarten students in bilingual classes, 27 Mexican-American students (Grades 1-2) in study previous school year, 23 Mexican-American students (Grades 1-2) added to the study C) Year Three--38 Black students (Grades 1-3) who spoke no Spanish, 56 Mexican-American children (Grades 1-3) previously in study, 81 Mexican-American students (Grades 1-3) new to the study D) Year Four--21 Black children (Grades 2-4) not previously in the program, 24 Black children (Grades 2-4) previously in program, 100

Mexican-American students (Grades 2-4) previously in program, 96 Mexican-American students (Grades 2-4) new to program.

E) Year Five--12 Black students (Grades 3-5) not previously in program, 17 Black students (Grades 3-5) previously in program, 116 Mexican-American students (Grades 3-5) previously in program, 53 Mexican-American students (Grades 3-5) not previously in program

F) Year Six--13 Black students (Grades 4-6) new to program, 18 Black students (Grades 4-6) previously in program, 79 Mexican-American students (Grades 4-6) new to program, 112 Mexican-American students (Grades 4-6) previously in program.

Control Group:

A) Year One--40 children in regular classroom (K-1) at same school

B) Year Two--unspecified number of children in regular classroom (some unspecified of which received English as a Second Language instruction)

C) Year Three--unspecified number of students in regular classes at a different school and unspecified number of students (one class per grade level 1-3) in target school

D) Year Four--unspecified number of children (one class for each of Grades 2-4) at target school

E) Year Five--unspecified number of students (one class each for Grades 3-5) at target school

F) Year Six--63 students in Grades 4-6 in regular classroom at target school

Duration: Six years (September, 1969-June, 1975).

Ages: Grades K-6

Type of Program: Transitional bilingual education

Description: Over the six-year period, attempts were made to evaluate the success of the Title VII Project in the Compton Unified School District. The design of the study changed from school year to school year, resulting in six different studies.

Originally, students during the first year of the program were placed into the bilingual classes on the results of the *Student Evaluation Scales* for listening, comprehension, and speaking, and the *Thomas Completion Stories* in Spanish. Students with the least facility in English received preference for placement. Pre-testing, consisting of the *Peabody Picture Vocabulary Test* and the *Stern Expressive Vocabulary Inventory*, was conducted during October, November, and December. The treatment group received bilingual instruction from Spanish-speaking teachers (no description of instruction was given), with the control group receiving instruction from English monolinguals (although Spanish-speaking aides were assigned to the control classes).

Post-test results (no statistics given) showed superior gains by experimental children; for example, kindergarten children scored 0 on the *Peabody Picture Vocabulary Test* on the pre-test, but averaged scores equivalent to seven-year-olds on the post-test. Other specific results were not given although "in terms of pre-/post-test gains, the bilingual classes were consistently superior to the controls."

During the second year the design was changed because of diffusion of ideas and materials between regular and bilingual teachers during the first year, which tended to contaminate the design. A control group at another school with a similar population was chosen; and the kindergarten, first- and second-grade classes at that school with the greatest number of non-English-speaking children were used. Children in the experimental groups continued to receive instruction from Spanish-speakers (except for one English monolingual who was replaced by a Spanish-speaking substitute for several months before a Spanish-speaking teacher who had no bilingual training was located). In the middle of the school year, it was discovered that some of the control group children received English as a Second Language instruction. The kindergarten and first-grade children in both groups were pre- and post-tested with the _Goodenough-Harris Draw-A-Man Test_, the _Expressive Vocabulary Inventory_, and the _Inter-American Oral Comprehension Test_, while the second-graders were administered the _Inter-American General Ability Test_ and _Reading Test_ in Spanish.

Throughout the school year, students who achieved what was considered to be a sufficient level of English abilities were transferred to regular classrooms, while new students with no English abilities were moved into the bilingual program to fill the vacancies. Despite this lack of consistency in the experimental groups, the treatment groups demonstrated significant gains superior to the controls in reading in Spanish and in oral English comprehension; their oral comprehension of Spanish was equal to that of the controls.

The third year of the project witnessed a major shift in the operation of the bilingual program. In order to meet changed federal guidelines, students in the bilingual program were mixed with students from other ethnic groups (primarily Black in this school district) to allay accusations of segregation in the bilingual classes. The permission of the parents of the Black children was required, so that selection was not random.

The instructional program also changed so that bilingual teachers became resource persons working with the children in the program within the regular Grades 1-3 classrooms and in conjunction with the regular classroom teachers. Because the control group from the year before had received English as a Second Language instruction, six new control classes (2 at each grade level 1-3) were chosen at a different school which resembled the target school in population. Additionally, comparison classes at the target school in Grades 1-3 were designated.

All students were pre- and post-tested with the Comprehensive Tests in Reading and Math; in addition, the experimental and control groups (not the comparison groups) were tested with the Inter-American Oral Comprehension, Reading, and General Ability Tests in Spanish. The covaried scores revealed no consistent pattern. For example, in reading, the Black and Chicano children in the experimental groups were significantly lower than any other group, while in the second grade, the Black control children and the Chicano experimental children scored the highest gains. In math, the Black and Chicano experimental students in the second grade showed significant gains; on the other hand, the Black and Chicano experimental youngsters in the first and third grades were at the lowest end of the distribution. On the Oral Comprehension Test in Spanish, the Chicano experimental children were consistently higher than any other group, while the Black treatment students in the first and third grades scored almost twice as high as their counterparts in control groups. On the Tests of General Ability in Spanish, the Chicano experimental youngsters scored higher than any other group; the Black experimental students scored on a comparable level to that of the Chicano controls.

During the third year, the Bilingual Affect Scale (no pilot, no reliability or validity established) was developed and administered. No differences were demonstrated for any groups from pre- to post-test.

At the beginning of the fourth year the idea of a formal control group was abandoned because of past difficulties. Children from the target school receiving regular instruction with socio-economic backgrounds and ethnic origins similar to the treatment children were selected in this and ensuing years to serve as comparisons. The experimental classes continued to operate with a team-teaching approach between the floating bilingual teachers and the regular classroom teachers served by them.

The scores on the Cooperative Tests in Reading and Math were again analyzed, and the progress of the experimental and comparison students was equivalent. The Inter-American General Ability and Reading Tests were given in Spanish to the experimental groups only. The second- and third-grade Chicano students showed significant gains in all subtests, with the Black students demonstrating inconsistent gains. At the fourth-grade level all children made significant gains on all subtests, except one requiring an understanding of relationships. During the fourth year attendance and library use data were also collected, but revealed no differences between groups of children.

The project design for the fifth year remained consistent with that of the fourth year; however, teacher turn-over was great; all teachers of the district went on strike during the fifth year, and bilingual program teachers and aides were informed that the program might not continue the next school year. These factors were purported by the author to have affected the gains by students in the bilingual program during 1973-74.

Fifth-year assessment was conducted on a pre-/post-test basis using the Comprehensive Tests in Reading and Math and the Inter-American General Ability and Reading Tests, in spite of the fact that the Inter-American Tests were not adequate in content and norming data for students at these high grade levels. Statistical tests showed the comparison groups in the fourth grade to be one year advanced from the experimental in English reading; all experimental classes showed very little gain, and all classes in the study fell below expected grade norms. The Inter-American Tests were analyzed in terms of children in the program for three or more years and students with less than three years of bilingual experience; in all cases students who had been in the program longer did not achieve as highly on the Spanish reading comprehension test as did those with less experience. The author speculated that this resulted from the fact that the students who remained in the program and were not transferred to regular classes had made slow progress in acquiring English language abilities and were probably slow learners.

In the final year of the project, the design remained the same as for the fourth and fifth years. The Inter-American Tests of General Ability and Reading were again used with the experimental students; not all students who were pre-tested were also post-tested. On the whole, relative to the general ability tests, the experimental groups all demonstrated significant gains on all of the subtests, especially at the fourth grade level. The measures of Spanish reading indicated normally expected grade-level gains; when comparing veteran students in the bilingual program to those with two years or less of experience, there was little difference between the two groups, except in abstract reasoning in which the more experienced youngsters scored significantly higher.

The California Test of Basic Skills was added to the battery, and the scores (pre-test scores had been covaried) for the bilingual children were significantly lower than those for the comparison students, even though the bilingual youngsters made significant gains over their pre-test scores. By skimming through the class lists, the author discovered that many of the highest scorers in the comparison groups had once been in the bilingual program.

In summing up the six years of the project, the author felt that the encouraging gains from the early years of the program were not maintained throughout the project, with evidence in the last year of the below-grade-level abilities of the bilingual children in English language skills. Although the Chicano treatment students demonstrated good progress in Spanish, "there was little success in achieving growth in the acquisition of Spanish by non-Spanish students in the bilingual classes."

Strengths:

A) There was an attempt to conduct a longitudinal study.

B) Mexican-American students were placed into the bilingual program on the basis of need as determined by measures of their English language facility.

C) The author consistently listed limitations of the study (lack of trained teachers, inappropriate control groups, etc.).

Discussion:

This study was plagued with design problems from the very beginning; at no time were students selected randomly. Throughout the history of this research, there were concurrent events which tended to contaminate the results; these included a high rate of teacher turn-over within the program, a teacher strike, funding reductions, the placement of Black students who spoke no Spanish into the experimental group, and the transfer of bilingual students successful in the acquisition of English into the regular classroom.

The constant assignment of new Mexican-American students into the experimental group made it impossible to observe the rigorous rules of research; the difficulties in obtaining the appropriate control groups led the author in the last three years of the study to abandon this procedure completely. Some stumbling blocks in test administration were encountered, such as the use of the Inter-American Tests with children of ages for whom the tests were not normed, or the hasty development of the Bilingual Affect Scale.

In mid-stream, the concept of bilingual treatment was changed from that of a self-contained bilingual class to a team-teaching approach. In addition, even in the beginning of the study, regular classroom and bilingual teachers shared materials, methods, and ideas, with the school district even paying for the Berlitz courses in Spanish for the regular classroom teachers. The lack of rigor in this study would seem to make the results suspect.

Reference: Zirkel, P.A. An evaluation of the effectiveness of selected experimental bilingual education programs in Connecticut. West Hartford, Connecticut: Hartford University, Connecticut Migratory Children's Program, May, 1972.

Baker de Kanter Review - p. 20-21, Chapter 2.

Available From: ERIC (ED 070 326); in microfiche collection of Perry-Castaneda Library, University of Texas.

Name of Study: An Evaluation of the Effectiveness of Selected Experimental Bilingual Education Programs in Connecticut.

Author and Date: Zirkel, Perry Alan; May, 1972.

Location: Bridgeport, Hartford, New Britain, and New London, Connecticut.

Treatment Group: Three treatment groups: A) 95 Puerto Rican, Spanish-speaking, economically disadvantaged students of Grades 1-3 in bilingual education classes. All subject matter was taught in Spanish with English as a Second Language provided; B) 15 Puerto Rican Spanish-dominant, economically disadvantaged students of Grades 1-3 who received tutoring assistance in the regular classroom in subject matter areas from bilingual aides. Iinstruction was 10 minutes to one hour which varied from school to school and included individual and small group instruction); C) 18 Puerto Rican, Spanish-dominant, economically disadvantaged students of Grades 1-3 who received 1 to 2 hours per day of resource assistance in content areas from bilingual teachers.

Control Group: Three control groups: A) 111 Puerto Rican, Spanish-dominant, economically disadvantaged students in Grades 1-3 in regular program with an ESL component; B) 13 Puerto Rican Spanish-dominant, economically disadvantaged students of Grades 1-3 in the regular program with an ESL component; C) 25 Puerto Rican, Spanish-dominant, economically disadvantaged students of Grades 1-3 in the regular program with an ESL component.

Duration: 1970-71 school year - (8 months pre- to post-test.)

Area: Grades 1-3.

Type of Program: Transitional bilingual education.

Description: The author sought to investigate 1) academic gains in English and Spanish made during the school year, 2) gains in self-concept, and 3) parental attitudes toward, interest in, and knowledge of the schools their children attended.

The experimental and control groups were matched as closely as possible for sex, age, language dominance, school, grade level, and socio-economic status (occupation of the head of the household). The tests used were the *Goodenough-Harris Draw-A-Man* (for intelligence data), the *Inter-American Tests of General Ability* (Spanish and English versions), the *Inferred Concept Scale* (for self-concept), and the *Zirkel-Greene Home Interview Schedule* (for parental attitudes). No indication was given as to how language dominance was established. These measures were administered in October and again in May, except for the *Home Interview* which was completed only in May.

Following pre-testing the author discovered that the scores on the *Tests of General Ability* and the self-concept scale were lower for the experimental than for the control group. Therefore, the IQ scores and the pre-test scores for the Tests of General Ability and the self-concept scale were covaried.

The result of the post-tests in May revealed that the bilingual classes all showed gains in general abilities in both English and Spanish and in self-concept. Significant gains occurred in the arena of self-concept for the first-graders in the bilingual classrooms and in the area of academics in both English and Spanish for the second- and third-graders in bilingual classrooms. The bilingual aide arrangement and the resource room concept also demonstrated more gains than the control group, although these gains were not statistically significant. The parent surveys which were completed with one of two families randomly selected from Hartford and Bridgeport did not reveal significant differences between the parents of children in the control group and those in the experimental group. The parents of the experimental group students tended to be more in favor of bilingual education, more involved, and more interested in schools.

Strengths:

A) The control groups were carefully chosen to resemble the experimental groups. Following pre-testing, students were dropped from the study to achieve matching for sex, age, socio-economic status, school, grade level, and language dominance.

B) The design used pre-tests and post-tests on both control and experimental groups, and the author covaried pre-test results on measures of IQ, general abilities, and self-concept.

C) The researcher visited all of the bilingual classes during the school year and administered a survey to the teachers to determine the actual models of instruction used in the bilingual classrooms. Classrooms which deviated from the ideal model of having most of the subject matter taught in Spanish with language instruction in English as a complement were screened into clear categories (aide program or resource room) or eliminated from the study.

Discussion: Although gains were shown for all groups receiving some sort of bilingual assistance, significant gains were found in two diverse groups. The author stated that significant gains in first-grade self-concept may have been the result of the impressionability of first-graders, and their lack of significant gains academically could be related to the fact that all first-grade teachers in the study were not as experienced or qualified (in the researcher's opinion) as the other teachers in the study.

The author stated that the results must be interpreted cautiously as the experimental program was in the first year of operation and the experimental mortality of all groups was about 25%. The study was not longitudinal, and the samples in most groups were relatively small.

The question of the establishment of language dominance is an open issue, as there was no reference as to what criteria were utilized to assign students to bilingual classrooms and to regular classrooms. The reader must also be reminded that all students in the study (experimental and control alike) received English as a Second Language instruction; therefore, the results of the control group reflect a variable not present in the regular classroom situation, which was not a part of the study.

"END OF DOCUMENT"